HAL
open science

Étude et calcul de quelques distances en probabilités et statistique et applications : séparation asymptotique des chaînes de Markov

Bernard Garel

▶ To cite this version:

Bernard Garel. Étude et calcul de quelques distances en probabilités et statistique et applications : séparation asymptotique des chaînes de Markov. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG; Université Joseph-Fourier - Grenoble I, 1983. tel-00308567

HAL Id: tel-00308567
https://theses.hal.science/tel-00308567v1

Submitted on 31 Jul 2008

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'Université Scientifique et Médicale de Grenoble

et à

l'Institut National Polytechnique de Grenoble

pour obtenir le grade de
DOCTEUR ES SCIENCES
«Mathématiques»

par

GAREL Bernard

ETUDE ET CALCUL DE QUELQUES DISTANCES EN PROBABILITES ET STATISTIQUE ET APPLICATIONS. SEPARATION ASYMPTOTIQUE DES CHAINES DE MARKOV.

Thèse soutenue le 28 juin 1983 devant la commission d'examen.

B. VAN CUTSEM	Président
J.R. BARRA	Examinateurs
D. BOSQ	
J. FONLUPT	
J. GEFFROY	
J.P. LAURENT	

Thèse préparée à l'Université de Savoie et au laboratoire IMAG.

UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

Année universitaire 1982-1983

Président de l'Université : M. TANCHE

MEMBRES DU CORPS ENSEIGNANT DE L'U.S.M.G.

(RANG A)

SAUF ENSEIGNANTS EN MEDECINE ET PHARMACIE

PROFESSEURS DE 1ère CLASSE

ARNAUD Paul	Chimie organique
ARVIEU Robert	Physique nucléaire I.S.N.
AUBERT Guy	Physique C.N.R.S.
AYANT Yves	Physique approfondie
BARBIER Marie-Jeanne	Electrochimie
BARBIER Jean-Claude	Physique expérimentale C.N.R.S. (labo de magnétisme)
BARJON Robert	Physique nucléaire I.S.N.
BARNOUD Fernand	Biosynthèse de la cellulose-Biologie
BARRA Jean-René	Statistiques - Mathématiques appliquées
BELORISKY Elie	Physique
BENZAKEN Claude (M.)	Mathématiques pures
BERNARD Alain	Mathématiques pures
BERTRANDIAS Françoise	Mathématiques pures
BERTRANDIAS Jean-Paul	Mathématiques pures
BILLET Jean	Géographie
BONNIER Jean-Marie	Chimie générale
BOUCHEZ Robert	Physique nucléaire I.S.N.
BRAVARD Yves	Géographie
CARLIER Georges	Biologie végétale
CAUQUIS Georges	Chimie organique
CHIBON Pierre	Biologie animale
COLIN DE VERDIERE Yves	Mathématiques pures
CRABBE Pierre (détaché)	C.E.R.M.O.
CYROT Michel	Physique du solide
DAUMAS Max	Géographie
DEBELMAS Jacques	Géologie générale
DEGRANGE Charles	Zoologie
DELOBEL Claude (M.)	M.I.A.G. Mathématiques appliquées
DEPORTES Charles	Chimie minérale
DESRE Pierre	Electrochimie
DOLIQUE Jean-Michel	Physique des plasmas
DUCROS Pierre	Cristallographie
FONTAINE Jean-Marc	Mathématiques pures
GAGNAIRE Didier	Chimie physique

.../...

GASTINEL Noël	Analyse numérique - Mathématiques appliquées
GERBER Robert	Mathématiques pures
GERMAIN Jean-Pierre	Mécanique
GIRAUD Pierre	Géologie
IDELMAN Simon	Physiologie animale
JANIN Bernard	Géographie
JOLY Jean-René	Mathématiques pures
JULLIEN Pierre	Mathématiques appliquées
KAHANE André (détaché DAFCO)	Physique
KAHANE Josette	Physique
KOSZUL Jean-Louis	Mathématiques pures
KRAKOWIAK Sacha	Mathématiques appliquées
KUPTA Yvon	Mathématiques pures
LACAZE Albert	Thermodynamique
LAJZEROWICZ Jeannine	Physique
LAJZEROWICZ Joseph	Physique
LAURENT Pierre	Mathématiques appliquées
DE LEIRIS Joël	Biologie
LLIBOUTRY Louis	Géophysique
LOISEAUX Jean-Marie	Sciences nucléaires I.S.N.
LOUP Jean	Géographie
MACHE Régis	Physiologie végétale
MAYNARD Roger	Physique du solide
MICHEL Robert	Minéralogie et pétrographie (géologie)
MOZIERES Philippe	Spectrométrie - Physique
OMONT Alain	Astrophysique
OZENDA Paul	Botanique (biologie végétale)
PAYAN Jean-Jacques (détaché)	Mathématiques pures
PEBAY PEYROULA Jean-Claude	Physique
PERRIAUX Jacques	Géologie
PERRIER Guy	Géophysique
PIERRARD Jean-Marie	Mécanique
RASSAT André	Chimie systématique
RENARD Michel	Thermodynamique
RICHARD Lucien	Biologie végétale
RINAUDO Marguerite	Chimie CERMAV
SENGEL Philippe	Biologie animale
SERGERAERT Francis	Mathématiques pures
SOUTIF Michel	Physique
VAILLANT François	Zoologie
VALENTIN Jacques	Physique nucléaire I.S.N.
VAN CUTSEN Bernard	Mathématiques appliquées
VAUQUOIS Bernard	Mathématiques appliquées
VIALON Pierre	Géologie

PROFESSEURS DE 2ème CLASSE

ADIBA Michel	Mathématiques pures
ARMAND Gilbert	Géographie

.../...

AURIAULT Jean-Louis	Mécanique
BEGUIN Claude (M.)	Chimie organique
BOEHLER Jean-Paul	Mécanique
BOITET Christian	Mathématiques appliquées
BORNAREL Jean	Physique
BRUN Gilbert	Biologie
CASTAING Bernard	Physique
CHARDON Michel	Géographie
COHENADDAD Jean-Pierre	Physique
DENEUVILLE Alain	Physique
DEPASSEL Roger	Mécanique des fluides
DOUCE Roland	Physiologie végétale
DUFRESNOY Alain	Mathématiques pures
GASPARD François	Physique
GAUTRON René	Chimie
GIDON Maurice	Géologie
GIGNOUX Claude (M.)	Sciences nucléaires I.S.N.
GUITTON Jacques	Chimie
HACQUES Gérard	Mathématiques appliquées
HERBIN Jacky	Géographie
HICTER Pierre	Chimie
JOSELEAU Jean-Paul	Biochimie
KERCKOVE Claude (M.)	Géologie
LE BRETON Alain	Mathématiques appliquées
LONGEQUEUE Nicole	Sciences nucléaires I.S.N.
LUCAS Robert	Physiques
LUNA Domingo	Mathématiques pures
MASCLE Georges	Géologie
NEMOZ Alain	Thermodynamique (CNRS - CRTBT)
OUDET Bruno	Mathématiques appliquées
PELMONT Jean	Biochimie
PERRIN Claude (M.)	Sciences nucléaires I.S.N.
PFISTER Jean-Claude (détaché)	Physique du solide
PIBOULE Michel	Géologie
PIERRE Jean-Louis	Chimie organique
RAYNAUD Hervé	Mathématiques appliquées
ROBERT Gilles	Mathématiques pures
ROBERT Jean-Bernard	Chimie physique
ROSSI André	Physiologie végétale
SAKAROVITCH Michel	Mathématiques appliquées
SARROT REYNAUD Jean	Géologie
SAXOD Raymond	Biologie animale
SOUTIF Jeanne	Physique
SCHOOL Pierre-Claude	Mathématiques appliquées
STUTZ Pierre	Mécanique
SUBRA Robert	Chimie
VIDAL Michel	Chimie organique
VIVIAN Robert	Géographie

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

Année universitaire 1982-1983

Président de l'Université : D. BLOCH

Vice-Président : René CARRE
Hervé CHERADAME
Marcel IVANES

PROFESSEURS DES UNIVERSITES :

ANCEAU François	E.N.S.I.M.A.G.
BARRAUD Alain	E.N.S.I.E.G.
BAUDELET Bernard	E.N.S.I.E.G.
BESSON Jean	E.N.S.E.E.G.
BLIMAN Samuel	E.N.S.E.R.G.
BLOCH Daniel	E.N.S.I.E.G.
BOIS Philippe	E.N.S.H.G.
BONNETAIN Lucien	E.N.S.E.E.G.
BONNIER Etienne	E.N.S.E.E.G.
BOUVARD Maurice	E.N.S.H.G.
BRISSONNEAU Pierre	E.N.S.I.E.G.
BUYLE BODIN Maurice	E.N.S.E.R.G.
CAVAIGNAC Jean-François	E.N.S.I.E.G.
CHARTIER Germain	E.N.S.I.E.G.
CHENEVIER Pierre	E.N.S.E.R.G.
CHERADAME Hervé	U.E.R.M.C.P.P.
CHERUY Arlette	E.N.S.I.E.G.
CHIAVERINA Jean	U.E.R.M.C.P.P.
COHEN Joseph	E.N.S.E.R.G.
COUMES André	E.N.S.E.R.G.
DURAND Francis	E.N.S.E.E.G.
DURAND Jean-Louis	E.N.S.I.E.G.
FELICI Noël	E.N.S.I.E.G.
FOULARD Claude	E.N.S.I.E.G.
GENTIL Pierre	E.N.S.E.R.G.
GUERIN Bernard	E.N.S.E.R.G.
GUYOT Pierre	E.N.S.E.E.G.
IVANES Marcel	E.N.S.I.E.G.
JAUSSAUD Pierre	E.N.S.I.E.G.
JOUBERT Jean-Claude	E.N.S.I.E.G.
JOURDAIN Geneviève	E.N.S.I.E.G.
LACOUME Jean-Louis	E.N.S.I.E.G.
LATOMBE Jean-Claude	E.N.S.I.M.A.G.

.../...

LESSIEUR Marcel	E.N.S.H.G.
LESPINARD Georges	E.N.S.H.G.
LONGEQUEUE Jean-Pierre	E.N.S.I.E.G.
MAZARE Guy	E.N.S.I.M.A.G.
MOREAU René	E.N.S.H.G.
MORET Roger	E.N.S.I.E.G.
MOSSIERE Jacques	E.N.S.I.M.A.G.
PARIAUD Jean-Charles	E.N.S.E.E.G.
PAUTHENET René	E.N.S.I.E.G.
PERRET René	E.N.S.I.E.G.
PERRET Robert	E.N.S.I.E.G.
PIAU Jean-Michel	E.N.S.H.G.
POLOUJADOFF Michel	E.N.S.I.E.G.
POUPOT Christian	E.N.S.E.R.G.
RAMEAU Jean-Jacques	E.N.S.E.E.G.
RENAUD Maurice	U.E.R.M.C.P.P.
ROBERT André	U.E.R.M.C.P.P.
ROBERT François	E.N.S.I.M.A.G.
SABONNADIERE Jean-Claude	E.N.S.I.E.G.
SAUCIER Gabrielle	E.N.S.I.M.A.G.
SCHLENKER Claire	E.N.S.I.E.G.
SCHLENKER Michel	E.N.S.I.E.G.
SERMET Pierre	E.N.S.E.R.G.
SILVY Jacques	U.E.R.M.C.P.P.
SOHM Jean-Claude	E.N.S.E.E.G.
SOUQUET Jean-Louis	E.N.S.E.E.G.
VEILLON Gérard	E.N.S.I.M.A.G.
ZADWORNY François	E.N.S.E.R.G.

PROFESSEURS ASSOCIES

BASTIN Georges	E.N.S.H.G.
BERRIL John	E.N.S.H.G.
CARREAU Pierre	E.N.S.H.G.
GANDINI Alessandro	U.E.R.M.C.P.P.
HAYASHI Hirashi	E.N.S.I.E.G.

PROFESSEURS UNIVERSITE DES SCIENCES SOCIALES (Grenoble II)

BOLLIET Louis
Chatelin Françoise

PROFESSEURS E.N.S. Mines de Saint-Etienne

RIEU Jean
SOUSTELLE Michel

CHERCHEURS DU C.N.R.S.

FRUCHART Robert	Directeur de Recherche
VACHAUD Georges	Directeur de Recherche

.../...

ALLIBERT Michel	Maître de Recherche
ANSARA Ibrahim	Maître de Recherche
ARMAND Michel	Maître de Recherche
BINDER Gilbert	
CARRE René	Maître de Recherche
DAVID René	Maître de Recherche
DEPORTES Jacques	
DRIOLE Jean	Maître de Recherche
GIGNOUX Damien	
GIVORD Dominique	
GUELIN Pierre	
HOPFINGER Emil	Maître de Recherche
JOUD Jean-Charles	Maître de Recherche
KAMARINOS Georges	Maître de Recherche
KLEITZ Michel	Maître de Recherche
LANDAU Ioan-Dore	Maître de Recherche
LASJAUNIAS J.C.	
MERMET Jean	Maître de Recherche
MUNIER Jacques	Maître de Recherche
PIAU Monique	
PORTESEIL Jean-Louis	
THOLENCE Jean-Louis	
VERDILLON André	

CHERCHEURS du MINISTERE de la RECHERCHE et de la TECHNOLOGIRE (Directeurs et Maîtres de Recherches, ENS Mines de St. Etienne)

LESBATS Pierre	Directeur de Recherche
BISCONDI Michel	Maître de Recherche
KOBYLANSKI André	Maître de Recherche
LE COZE Jean	Maître de Recherche
LALAUZE René	Maître de Recherche
LANCELOT Francis	Maître de Recherche
THEVENOT François	Maître de Recherche
TRAN MINH Canh	Maître de Recherche

PERSONNALITES HABILITEES à DIRIGER des TRAVAUX de RECHERCHE (Décision du Conseil Scientifique)

ALLIBERT Colette	E.N.S.E.E.G.
BERNARD Claude	E.N.S.E.E.G.
BONNET Rolland	E.N.S.E.E.G.
CAILLET Marcel	E.N.S.E.E.G.
CHATILLON Catherine	E.N.S.E.E.G.
CHATILLON Christian	E.N.S.E.E.G.
COULON Michel	E.N.S.E.E.G.
DIARD Jean-Paul	E.N.S.E.E.G.
EUSTAPOPOULOS Nicolas	E.N.S.E.E.G.
FOSTER Panayotis	E.N.S.E.E.G.

.../...

GALERIE Alain	E.N.S.E.E.G.
HAMMOU Abdelkader	E.N.S.E.E.G.
MALMEJAC Yves	E.N.S.E.E.G. (CENG)
MARTIN GARIN Régina	E.N.S.E.E.G.
NGUYEN TRUONG Bernadette	E.N.S.E.E.G.
RAVAINE Denis	E.N.S.E.E.G.
SAINFORT	E.N.S.E.E.G. (CENG)
SARRAZIN Pierre	E.N.S.E.E.G.
SIMON Jean-Paul	E.N.S.E.E.G.
TOUZAIN Philippe	E.N.S.E.E.G.
URBAIN Georges	E.N.S.E.E.G. (Laboratoire des ultra-réfractaires ODEILLON)
GUILHOT Bernard	E.N.S. Mines Saint Etienne
THOMAS Gérard	E.N.S. Mines Saint Etienne
DRIVER Julien	E.N.S. Mines Saint Etienne
BARIBAUD Michel	E.N.S.E.R.G.
BOREL Joseph	E.N.S.E.R.G.
CHOVET Alain	E.N.S.E.R.G.
CHEHIKIAN Alain	E.N.S.E.R.G.
DOLMAZON Jean-Marc	E.N.S.E.R.G.
HERAULT Jeanny	E.N.S.E.R.G.
MONLLOR Christian	E.N.S.E.R.G.
BORNARD Guy	E.N.S.I.E.G.
DESCHIZEAU Pierre	E.N.S.I.E.G.
GLANGEAUD François	E.N.S.I.E.G.
KOFMAN Walter	E.N.S.I.E.G.
LEJEUNE Gérard	E.N.S.I.E.G.
MAZUER Jean	E.N.S.I.E.G.
PERARD Jacques	E.N.S.I.E.G.
REINISCH Raymond	E.N.S.I.E.G.
ALEMANY Antoine	E.N.S.H.G.
BOIS Daniel	E.N.S.H.G.
DARVE Félix	E.N.S.H.G.
MICHEL Jean-Marie	E.N.S.H.G.
OBLED Charles	E.N.S.H.G.
ROWE Alain	E.N.S.H.G.
VAUCLIN Michel	E.N.S.H.G.
WACK Bernard	E.N.S.H.G.
BERT Didier	E.N.S.I.M.A.G.
CALMET Jacques	E.N.S.I.M.A.G.
COURTIN Jacques	E.N.S.I.M.A.G.
COURTOIS Bernard	E.N.S.I.M.A.G.
DELLA DORA Jean	E.N.S.I.M.A.G.
FONLUPT Jean	E.N.S.I.M.A.G.
SIFAKIS Joseph	E.N.S.I.M.A.G.
CHARUEL Robert	U.E.R.M.C.P.P.
CADET Jean	C.E.N.G.
COEURE Philippe	C.E.N.G. (LETI)

.../...

DELHAYE Jean-Marc	C.E.N.G. (STT)
DUPUY Michel	C.E.N.G. (LETI)
JOUVE Hubert	C.E.N.G. (LETI)
NICOLAU Yvan	C.E.N.G. (LETI)
NIFENECKER Hervé	C.E.N.G.
PERROUD Paul	C.E.N.G.
PEUZIN Jean-Claude	C.E.N.G. (LETI)
TAIEB Maurice	C.E.N.G.
VINCENDON Marc	C.E.N.G.

LABORATOIRES EXTERIEURS

DEMOULIN Eric	C.N.E.T.
DEVINE	C.N.E.T. (R.A.B.)
GERBER Roland	C.N.E.T.
MERCKEL Gérard	C.N.E.T.
PAULEAU Yves	C.N.E.T.
GAUBERT C.	I.N.S.A. Lyon

Monsieur le Professeur VAN CUTSEM m'a accueilli il y a dix ans dans son laboratoire et m'a aidé à parcourir ce long chemin. Monsieur le Professeur GEFFROY m'a donné les moyens de terminer ce travail et les motivations pour y parvenir. Qu'ils trouvent ici l'expression de ma reconnaissance.

Que Messieurs les Professeurs BARRA, BOSQ, FONLUPT, LAURENT soient vivement remerciés de l'honneur qu'ils me font en participant au jury de cette thèse.

Que tous ceux qui m'ont aidé acceptent mes remerciements, en particulier : Erik GEISSLER qui, à Chambéry, m'a donné l'indépendance administrative et les moyens financiers nécessaires à la réalisation de ce travail , Christophe LACOTE qui a programmé l'algorithme de Ford et Fulkerson , enfin les personnes qui ont participé à la frappe et au tirage de cette thèse.

TABLE DES MATIERES

INTRODUCTION 1

CHAPITRE I

I - 1 La distance de Prokhorov

- Introduction, définition et quelques résultats généraux..... 8
- Calcul de la distance de Prokhorov 16
- Quelques résultats complémentaires 41
- Bibliographie du paragraphe I - 1..... 48

I - 2 La distance de Geffroy

- Définition, quelques propriétés et calcul de la distance de Geffroy 50

I - 3 La distance de Fortet-Mourier-Wasseirstein.

I - 3 - 1 Introduction 59

Généralités 60

I - 3 - 2 Calcul de la distance de Fortet-Mourier 68

I - 3 - 3 Exemples sur $\mathbb{R}$ et $\mathbb{R}^2$ 78

I - 3 - 4 Un aspect du théorème de Kantorovitch et Rubinstein 89

Bibliographie du paragraphe I - 3 106

CHAPITRE II

II - 1 Un contre-exemple pour l'étude de la normalité asymptotique de l'estimateur ensembliste du maximum de vraisemblance..... 108
- cas non régulier -

II - 2 Visualisation d'échantillons multidimensionnels en vue de la détection de valeurs aberrantes 117

II - 3 Similarité des estimateurs 130

Bibliographie du chapitre II 142

CHAPITRE III

III - 1 Introduction 144

III - 2 Analyse spectrale du problème 145

III - 3 Séparation asymptotique des chaînes de Markov 148

III - 4 Problèmes liés à l'estimation du paramètre θ..... 160

Appendice 166

Bibliographie du chapitre III 169

INTRODUCTION

Ce travail comporte trois parties. Le chapitre I est consacré à l'étude de distances qui interviennent en Probabilités ou Statistique et qui ont rapport avec la topologie de la convergence faible (ou étroite) des probabilités. Il s'agit principalement de la distance de Prokhorov, de la distance de Geffroy et de la distance de Wasserstein.

La première de ces distances a été introduite par Prokhorov en 1956. Soit $(\mathcal{X}, d)$ un espace polonais, $\mathcal{B}$ sa tribu des boréliens et P et Q deux probabilités sur $(\mathcal{X}, \mathcal{B})$. La distance de Prokhorov notée $\Pi(P, Q)$ est définie par : $\Pi(P, Q) = \inf \{ \varepsilon \,/\, \forall A \text{ fermé } \in \mathcal{B} : P(A) \leq Q(A^{\varepsilon}) + \varepsilon \}$ où pour tout $\varepsilon > 0$ A^{ε} est le dilaté d'ordre ε de A.

Prokhorov a introduit cette distance pour l'étude de problèmes de convergence de processus stochastiques. Il a montré que la convergence au sens de cette distance est équivalente à la convergence étroite des mesures. En 1965, Strassen a permis de développer l'utilisation de cette distance grâce au théorème qui porte son nom et qui lie la valeur de la distance $\Pi(P, Q)$ à la charge d'une bande diagonale de l'espace $\mathcal{X} \times \mathcal{X}$ par une mesure λ dont les marginales sont P et Q.

En 1968, Hampel a proposé d'utiliser cette distance pour définir une

notion qualitative de robustesse : le premier ε de la définition apparaissant dans le terme A^{ε} et précisant la dilatation de A, peut en effet symboliser des erreurs d'arrondi et le second ε : des erreurs grossières.

La distance de Prokhorov a donc pris de plus en plus d'importance tant en Probabilités qu'en Statistique. Mais elle reste d'une perception non triviale. Elle tient compte à la fois de l'éloignement du support des mesures au sens d'une métrique choisie, et de leurs différences en tant que mesures. Mis à part des cas très simples, on ne disposait pas de méthodes pour la calculer.

On démontre tout d'abord diverses variantes de la définition. Puis on s'attaque plus spécifiquement au calcul de Π (P, Q). On résoud ce problème par approximation : on construit deux suites (P_n) et (Q_n) de probabilités à support fini qui convergent étroitement vers P et Q. Puis on approche Π (P,Q) par Π (P_n, Q_n).

Le calcul de Π (P_n, Q_n) sur ordinateur revient à la recherche d'un flot maximum sur un réseau, à l'aide de l'algorithme de Ford et Fulkerson. On traite alors divers exemples, avant de démontrer quelques résultats complémentaires permettant la simplification éventuelle de certains calculs de Π(P, Q).

Dans un second paragraphe, on étudie la distance de Geffroy. Introduite en 1973, cette distance s'applique aussi bien aux répartitions ponctuelles qu'aux mesures positives, finies sur tout borélien borné. Soit $(\mathcal{X}, ||.||)$ un espace vectoriel séparable de norme $||.||$. Soient f et g deux répartitions ponctuelles sur $\mathcal{X}$; N (f,A) l'effectif de f sur $A \subset \mathcal{X}$, $r \in \mathbb{R}^{+*}$, b(o,r) la boule

ouverte de centre o de rayon r, on définit :

$$\sigma_r\ (f,g) = \inf \{\delta > o / \forall A \in \mathcal{B} \text{ tel que}$$

$$A^\delta \in b\ (o,r) : N(f,A) \leq N(g,A^\delta) + \delta \}$$

puis $\rho_r\ (f,g) = \max\ (\sigma_r\ (f,g)\ ,\ \sigma_r\ (g,f))$

enfin $$\rho\ (f,g) = \sum_{m=1}^{+\infty} 2^{-m} \frac{\rho_m\ (f,g)}{1 + \rho_m\ (f,g)}$$

qui est la distance de Geffroy entre f et g.

On montre que cette quantité est comprise entre 0 et 2 (1-Log 2). Puis on s'intéresse à diverses équivalences et au calcul de cette distance qui se fera en général de façon approchée.

Le problème de la recherche de la distance sur les variables aléatoires dont la distance de Geffroy est la distance minimale suivant le concept de Zolatarev, reste ouvert.

Poursuivant notre étude, on arrive alors à la distance définie par Fortet et Mourier en 1953, alors qu'ils étudiaient le problème de la convergence de la répartition empirique vers la distribution théorique. Reprenons l'espace polonais du début. Soit une fonction $f : \mathcal{X} \to \mathbb{R}$

On définit :

$$\| f \|_L = \sup_{x \neq y} \frac{|f(x) - (f(y)|}{d(x,y)}$$

$$\| f \|_\infty = \sup_x | f(x) |$$

$$\| f \|_{BL} = \| f \|_L + \| f \|_\infty$$

la première étant une semi-norme et les deux autres des normes. Alors pour deux probabilités P et Q (et plus généralement pour deux mesures) vérifiant une con-

dition relative à la distance d on pose :

$$||P-Q||_L^* = \sup \left\{ \left| \int f \, d(P-Q) \right| / ||f||_L \leq 1 \right\}$$

qu'on appellera distance de Fortet-Mourier.

Dudley a étudié la variante importante suivante :

$$d_{BL}(P,Q) = \sup \left\{ \left| \int f \, d(P-Q) \right| / ||f||_{BL} \leq 1 \right\}$$

La convergence ou sens de cette dernière distance est équivalente à la convergence étroite des mesures et d_{BL} est elle-même équivalente à $\Pi(P,Q)$.

P.J. Huber a lui aussi proposé une variante en supposant que d est comprise entre 0 et 1, quitte à la remplacer le cas échéant par $\frac{d}{1+d}$.

Après quelques rappels sur l'algèbre des fonctions lipschitziennes et sur des résultats de densité que les propriétés énoncées ci-dessus laissent entrevoir, on s'intéresse au calcul de la distance de Fortet-Mourier. On ramène ce problème, par discrétisation, à un nouveau problème de recherche opérationnelle : le problème de Hitchcock.

On définit alors une dernière distance : la distance de Wasserstein. Si P et Q sont deux probabilités sur $(\mathcal{X}, d)$:

$$W(P,Q) = \inf \left\{ \int d(x,y) \, d\mu \; (x,y) / \mu \circ \Pi_1^{-1} = P, \; \mu \circ \Pi_2^{-1} = Q \right\}$$

où Π_i est la $i^{\text{ème}}$ projection.

Il se trouve que la distance de Wasserstein est égale à la distance

de Fortet-Mourier, résultat connu sous le nom de théorème de Kantorovitch et Rubinstein, dont une démonstration dans le cas d'un espace polonais, vient d'être publiée.

On introduit alors l'espace probabilisé $([0,1], \mathcal{B}([0,1], \lambda)$ où λ est la mesure de Lebesgue. On considère les variables aléatoires Y et Z définies sur $([0,1], \mathcal{B}[0,1], \lambda)$ à valeurs dans $(\mathcal{X}, d)$ de lois P et Q. Alors on peut écrire :

$$W(P,Q) = \inf \left\{ E(d(Y,Z)) \,/\, Y \text{ de loi } P \;,\; Z \text{ de loi } Q \right\}$$

On cherche à savoir si l'infimum est un minimum. Plus précisément si P et Q sont à support fini, on montre que pour toute variable aléatoire Y de loi P, il existe une variable aléatoire Z de loi Q telle que

$$E(d(Y,Z)) = \|P-Q\|_L^*$$

On étudie alors ce que devient ce résultat dans le cas continu. On montre que d'une façon générale le résultat ci-dessus n'est plus vrai, mais l'est très souvent lorsque $\mathcal{X} = \mathbb{R}$.

Le chapitre II est consacré aux problèmes d'estimation. On s'intéresse d'abord aux propriétés asymptotiques. On montre à l'aide d'un contre-exemple que la notion classique de normalité asymptotique de l'estimateur du maximum de vraisemblance ne se généralise pas de façon canonique au cas de l'estimateur ensembliste du maximum de vraisemblance, et de non régularité.

Puis on s'intéresse aux problèmes de fiabilité d'un estimateur relativement à l'instabilité éventuelle des modèles. On sait que les hypothèses sur les modèles statistiques sont rarement vérifiées. Ce qui veut dire qu'il y a

rarement équidistribution - souvent supposée gaussienne - ou indépendance. Les observations proviennent de lois mélangées. Parfois la variabilité intrinsèque des données entraîne elle aussi la présence de valeurs extrêmes. Après un rappel de la méthodologie statistique en matière de valeurs aberrantes on s'intéresse à la détection de ces valeurs dans le cas multidimensionnel, par visualisation. On montre que cette méthode ne peut être qu'une étape de la saisie et du tri éventuel des données.

On termine par une brève introduction à la robustesse. On rappelle alors que c'est plus de soixante-dix estimateurs de position qui ont été recensés lors de l'étude de robustesse de Princeton. On tente alors de procéder à un classement de ces estimateurs. On retient un certain nombre d'échantillons connus sur lesquels on fait opérer divers estimateurs. A l'aide d'une analyse en composantes principales, on montre qu'un facteur important (près de 60% de l'inertie totale) de séparation est le poids qu'accorde chaque estimateur aux valeurs extrêmes.

Le chapitre III nous donne l'occasion de reparler de distances entre probabilités. Puisqu'il se trouve que c'est en travaillant sur les chaines de Markov que Wasserstein a introduit sa distance et que la notion de séparation asymptotique est définie à l'aide des distances en variation totale et de Hellinger. Si (X_n) et (Y_n) sont deux chaines de Markov homogènes de même loi initiale et de matrices de transition respectives P et Q, on dira qu'il y a séparation asymptotique des deux chaînes si et seulement si :

$$\lim_{n \to +\infty} d_V \left(P^{(n)}, Q^{(n)}\right) = 1$$

où d_V est la distance en variation totale et $P^{(n)}$ (resp. $Q^{(n)}$) la loi conjointe de $(X_1, \ldots, X_n)$ (resp. de $(Y_1, \ldots, Y_n)$).

Après une étude matricielle, on obtient le résultat suivant dans le cas d'un espace d'états de cardinal fini :

Une condition nécessaire et suffisante pour qu'il n'y ait pas séparation asymptotique est qu'il existe une classe finale commune à P et Q telle que les restrictions respectives de P et Q à cette classe finale soient égales.

On suppose alors que les éléments p_{ij} des matrices de transition sont fonction d'un paramètre $\theta \in \mathbb{R}^r$ et l'on énonce une condition pour que pour tout θ, l'ensemble des θ' pour lesquels il n'y a pas séparation asymptotique soit de mesure de Lebesgue sur $\mathbb{R}^r$ nulle.

CHAPITRE I

Ce chapitre est consacré à l'étude de trois distances entre mesures, qui ont un rapport plus ou moins direct avec la convergence faible (ou étroite) des probabilités. On étudie successivement la distance de Prokhorov, une distance entre répartitions ponctuelles, introduite par Geffroy, enfin une distance issue d'une norme qui semble avoir été introduite par Fortet et Mourier et qui est égale à une autre distance couramment appelée distance de Wasserstein. On s'intéresse en particulier au calcul de ces distances, problème qui sera résolu de façon approchée avec l'aide d'un ordinateur. On précise certains liens de ces distances avec d'autres distances entre probabilités ou entre variables aléatoires de lois correspondantes.

I - 1. La distance de Prokhorov

En 1956, Prokhorov a introduit cette distance pour l'étude de problèmes de convergence de processus stochastiques. Il a tout de suite montré que la topologie associée était celle de la convergence étroite des mesures et l'on aurait pu croire que le plus important était dit. Mais en 1965, Strassen a démontré un résultat théorique d'utilisation courante actuellement.

Puis en 1968, Hampel a proposé d'utiliser cette distance pour la notion qualitative de robustesse.

Tant et si bien que cette distance a pris de plus en plus d'importance en probabilité comme en statistique, importance que la dernière décennie a confirmée.

D'une perception non triviale, elle tient compte à la fois de l'éloignement du support des mesures ou sens d'une métrique choisie préalablement et de leurs différences en tant que mesures. La distance de Prokhorov dépend donc d'une première métrique sur l'espace des événements élémentaires. On appellera $\Pi(P,Q)$ladistance de Prokhorov entre les probabilités P et Q. On part de la définition la plus couramment fournie de la distance de Prokhorov, on donne quelques équivalences puis on s'attaque au calcul de $\Pi(P, Q)$. Après avoir constaté qu'il n'est pas possible d'obtenir un résultat général pour le calcul exact, on résoud ce problème par approximation.

On construit deux suites (P_n) et $Q_n)$ de probabilités à support fini qui convergent étroitement vers P et Q puis on approche Π (P, Q) par Π (P_n, Q_n). On démontre enfin quelques résultats complémentaires.

Soit (S,d) un espace métrique séparable et complet. On appellera $\mathcal{B}$ la tribu des boréliens de S . Soit A un sous-ensemble non-vide de S et $\epsilon > 0$. On définit respectivement : $A^\epsilon = \{x \in S / \exists a \in A$ avec $d(x,a) < \epsilon\}$ et $A^{\epsilon]} = \{x \in S / \exists a \in A$ avec $d(x,a) \leq \epsilon\}$. Si A est vide, nous posons $A^\epsilon = \phi$. Alors, (cf. [3], page 113) $A^\epsilon = \bigcup_{a \in A} B(a,\epsilon)$ où $B(a,\epsilon)$ est la boule ouverte de centre a de rayon ϵ . Donc A^ϵ est un ouvert. D'autre part, (cf. [11], page 7) $A^{\epsilon]}$ est fermé. On a en effet les relations : si $\hat{A}$ désigne l'adhérence de A , $A^{\epsilon]} = \widehat{A^{\epsilon]}} = \hat{A}^{\epsilon]} = \widehat{\hat{A}^{\epsilon]}}$. On vérifie facilement que pour tout couple $(\epsilon_1, \epsilon_2) \in \mathbb{R}^{+*} \times \mathbb{R}^{+*}$, $(A^{\epsilon_1})^{\epsilon_2} \subset A^{\epsilon_1 + \epsilon_2}$. Enfin, désignons par $\bar{A}$ le complémentaire de A

1. - LEMME. <u>Soit</u> A <u>un sous-ensemble non vide de</u> S . <u>On a la relation</u>

$\overline{A^\epsilon}^\epsilon \subset \bar{A}$.

<u>De plus, l'inclusion est stricte si</u> A <u>est un ouvert</u> $\neq S$.

<u>Preuve</u>. - Si $\overline{A^\epsilon}$ est vide, la relation est triviale. Dans le cas contraire, soit $x \in \overline{A^\epsilon}^\epsilon$: $\exists a_o \in \overline{A^\epsilon}$ tel que

$$d(x,a_o) < \epsilon . \text{Or} (a_o \notin A^\epsilon) \Rightarrow (\forall a \in A \ , \ d(a,a_o) > \epsilon) \ .$$

De l'inégalité $d(x,a) > d(a,a_o) - d(x,a_o)$ on déduit $d(x,a) > 0$, soit $x \in \bar{A}$.

Pour conclure la preuve de la proposition, remarquons que si A est ouvert $\bar{A}$ est fermé, alors que $\overline{A^\epsilon}^\epsilon$ est ouvert.

2. - DEFINITION. <u>Soient</u> P <u>et</u> Q <u>deux probabilités sur</u> $(S,\mathcal{B})$. <u>On définit le réel</u> > 0 : $\Pi(P,Q)$ <u>par</u>

$$\Pi(P,Q) = \inf\{\epsilon / \forall A \in \mathcal{B} \quad P(A) \leq Q(A^{\epsilon}) + \epsilon\} .$$

3. - <u>Remarque</u>. L'application $\epsilon \to Q(A^{\epsilon}) + \epsilon$ est strictement croissante et l'on a :

$$(\Pi(P,Q) \leq \eta) \Leftrightarrow \forall \epsilon > \eta \ , \ \forall A \in \mathcal{B} \ , \ P(A) < Q(A^{\epsilon}) + \epsilon \ .$$

4. - <u>Remarque</u>. Dans la définition 2, on peut remplacer A^{ϵ} par $A^{\epsilon]}$ sans changer la valeur de $\Pi(P,Q)$. En effet, si l'on pose

$$\Pi_1(P,Q) = \inf\left\{\epsilon / \forall A \in \mathcal{B}, \ P(A) \leq \left(A^{\epsilon]}\right) + \epsilon\right\} ,$$

on a :

$$P(A) \leq Q(A^{\epsilon}) + \epsilon \leq Q\left(A^{\epsilon]}\right) + \epsilon \ ,$$

donc :

$$\Pi_1 \leq \Pi \ .$$

Inversement, supposons $\Pi > \Pi_1$ et soient ϵ_1 et ϵ_2 tels que :

$$\Pi > \epsilon_2 > \epsilon_1 > \Pi_1$$

$$\left\{\begin{array}{l} \epsilon_1 > \Pi_1(P,Q) \Rightarrow \forall A \in \mathcal{B} \quad P(A) < Q\left(A^{\epsilon_1]}\right) + \epsilon_1 \\ \text{et} \quad \epsilon_2 < \Pi(P,Q) \Rightarrow \exists A_o \in \mathcal{B} \ \text{ tel que } \ P(A_o) > Q\left(A_o^{\epsilon_2}\right) + \epsilon_2 \end{array}\right.$$

$$\Rightarrow \quad Q\left(A_o^{\epsilon_2}\right) + \epsilon_2 < P(A_o) < Q\left(A_o^{\epsilon_1]}\right) + \epsilon_1$$

mais $\quad A_o^{\epsilon_1]} \subset A_o^{\epsilon_2} \Rightarrow Q\left(A_o^{\epsilon_1]}\right) \leq Q\left(A_o^{\epsilon_2}\right)$

soit en remplaçant dans l'inégalité précédente :

$$Q\left(A_o^{\epsilon_1]}\right) + \epsilon_2 < Q\left(A_o^{\epsilon_1]}\right) + \epsilon_1$$

donc

$$\epsilon_2 < \epsilon_1$$

ce qui contredit l'hypothèse. Donc,

$$\Pi \leq \Pi_1 \ .$$

On en conclut que :

$$\Pi = \Pi_1 \ .$$

5. - Remarque. Dans la définition 2, on peut remplacer "$\forall A \in \mathcal{F}$" par "$\forall A$ fermé $\in \mathcal{F}$" sans changer la valeur de $\Pi(P,Q)$. En effet, si l'inégalité est vraie pour tout A , elle est vraie pour A fermé ! Inversement, supposons que $\forall A$ fermé $\in \mathcal{F}$, $P(A) \leq Q(A^{\epsilon}) + \epsilon$, alors soit B un borélien quelconque : on a $P(B) \leq P(\widehat{B}) \leq Q(\widehat{B}^{\epsilon}) + \epsilon = Q(B^{\epsilon}) + \epsilon$, car $\widehat{B}^{\epsilon} = B^{\epsilon}$.

6. - Remarque. Les remarques 4 et 5 donnent lieu à plusieurs variantes de la définition 2.

7. - PROPOSITION. Π est une distance sur l'espace des probabilités sur $(S,\mathcal{B})$. Elle prend ses valeurs dans $[0,1]$.

Preuve. -

1) Montrons tout d'abord la symétrie. Soit $\epsilon > 0$ et supposons que $\forall A \in \mathcal{B}$, $P(A) \leq Q(A^{\epsilon}) + \epsilon$. Alors pour un A quelconque, on a $P(\overline{A^{\epsilon}}) \leq Q(\overline{A^{\epsilon}}^{\epsilon}) + \epsilon \leq Q(\bar{A}) + \epsilon$, la dernière inégalité provenant du lemme 1. D'où $1 - P(A^{\epsilon}) \leq 1 - Q(A) + \epsilon$ soit $Q(A) \leq P(A^{\epsilon}) + \epsilon$. On en conclut que $\Pi(P,Q) = \Pi(Q,P)$.

2) Montrons maintenant que $\Pi(P,Q) = 0 \Leftrightarrow P \equiv Q$. Si $P = Q$, on a $\forall A \in \mathcal{B}$, $P(A) = Q(A) \Rightarrow \Pi(P,Q) = 0$. Réciproquement, soit B un fermé de $\mathcal{B}$. Pour tout $\epsilon > 0$, d'après la remarque 3, on a : $P(B) \leq Q(B^{\epsilon}) + \epsilon$. Comme B est fermé $\lim_{\epsilon \to 0} G(B^{\epsilon}) = G(B)$, d'où $P(B) \leq Q(B)$. Par symétrie, on obtient $Q(B) \leq P(B)$; soit pour tout fermé B : $P(B) = Q(B)$. On en conclut (cf. [13] page 162, propriété (μ.5)) que $F \equiv G$.

3) Supposons enfin que $\Pi(P,Q) = \epsilon_1$ et $\Pi(Q,T) = \epsilon_2$ et soit $\epsilon > \epsilon_1 + \epsilon_2$. On peut écrire $\epsilon = \epsilon'_1 + \epsilon'_2$ avec $\epsilon'_1 > \epsilon_1$ et $\epsilon'_2 > \epsilon_2$. D'après la remarque 3, on a :

$$\forall A \in \mathcal{B} , \quad P(A) < Q\left(A^{\epsilon'_1}\right) + \epsilon'_1 \quad \text{et} \quad Q(A) < T\left(A^{\epsilon'_2}\right) + \epsilon'_2$$

$$\Rightarrow P(A) < T\left[\left(A^{\epsilon'_1}\right)^{\epsilon'_2}\right] + \epsilon'_1 + \epsilon'_2 \leq T(A^{\epsilon}) + \epsilon .$$

On en conclut que $\Pi(P,T) \leq \epsilon_1 + \epsilon_2 = \Pi(P,Q) + \Pi(Q,T)$.

8. - <u>Autres variantes.</u> Soit A un élément quelconque de $\mathcal{B}$. Posons :

$$\epsilon(A) = \inf\{\eta / P(A) \leq Q(A^{\eta}) + \eta\} ,$$

alors (cf. [22] ou [16])

(9) $$\Pi(P,Q) = \sup\{\epsilon(A) / A \in \mathcal{B}\} .$$

Zolotarev dans [22] a en fait écrit :

$$\Pi(P,Q) = \inf_{\epsilon>0} \sup_{A} \max[\epsilon , P(A) - Q(A^{\epsilon})] .$$

Soit $\alpha > 0$. Posons $\beta(\alpha) = \inf\{\beta / \forall A \in \mathcal{B} , P(A) \leq Q(A^{\alpha}) + \beta\}$, alors (cf. [5] ou [18])

(10) $$\Pi(P,Q) = \inf\{\alpha / \beta(\alpha) \leq \alpha\} .$$

11. - <u>Distance de Prokhorov, produit de convolution et fonction caractéristique.</u>

Plusieurs majorations de la distance de Prokhorov ont été récemment fournies par des chercheurs soviétiques. Il s'agit de résultats très techniques faisant intervenir à la fois le produit de convolution et les fonctions caractéristiques. Nous rapportons ci-dessous les principaux d'entre eux. L'intérêt de ces résultats est qu'ils s'appliquent dans le cas multidimensionnel. Par contre, les auteurs n'ont donné aucun exemple, ni l'ordre de grandeur des erreurs d'approximation. Pour les démonstrations, nous renvoyons aux articles concernés.

Soient P, Q, S des lois de probabilité sur $\mathbb{R}^p$ muni d'une norme noté $\|.\|$.

La distance de Prokhorov est régulière relativement au produit de convolution sur $\mathbb{R}^p$ et si l'on note δ_0 la masse de Dirac en 0 , on a :

$$\Pi(P,Q) \leq \Pi(P*S\ ,\ Q*S) + 2\Pi(\delta_0, S)$$

(cf. [22]).

Yurinskii [21] propose tout d'abord sans hypothèse particulières l'inégalité suivante :

$$(12) \qquad \Pi(P,Q) \leq \frac{\Pi(P*S\ ,\ Q*S) + 3\max(\delta, S\{x/\|x\| \geq \delta\})}{S\{x/\|x\| < \delta\}} .$$

Puis, si l'on suppose que Q admet une densité $g(x)$ vérifiant :

$$(13) \qquad \int_{\mathbb{R}^p} \|g(x+h) - g(x)\| dx \leq \Gamma.\|h\| ,$$

on a :

$$(14) \qquad \Pi(P,Q) \leq C_1(1+\Gamma)\int_{\mathbb{R}^p} \|x\|\, S(dx) + C_2\Pi(P*S\ ,\ Q*S)$$

où C_1 et C_2 sont deux constantes qui ne dépendent que de p et de la norme $\|.\|$.

Si l'on note $r = [\frac{p}{2}] + 1$, si l'on appelle φ et ψ les fonctions caractéristiques des lois de probabilité P et Q et en supposant que P et Q ont des moments d'ordre 2 , soit :

$$(15) \qquad \int_{\mathbb{R}^p} \|x\|^2 P(dx) < +\infty \quad \text{et} \quad \int_{\mathbb{R}^p} \|x\|^2 Q(x) < +\infty ,$$

alors :

$$(16) \qquad \Pi(P,Q) \leq C\left\{\frac{1+\Gamma}{T} + \left(\int_{\|t\| \leq T} \sum_{(u) \leq r} \|D^u \Delta(t)\|^2 dt\right)^{\frac{1}{2}}\right\}$$

où $\quad u = u_1, \ldots, u_p \qquad (u) = u_1 + \ldots + u_p$

et $\quad D^u = \dfrac{\partial^{(u)}}{\partial t_1^{u_1} \ldots \partial t_p^{u_p}} \qquad \Delta(t) = \varphi(t) - \psi(t)$.

Cette dernière inégalité est alors utilisée pour l'étude de vitesse de convergence vers la loi normale, problème sur lequel nous reviendrons ultérieurement.

Abramov a amélioré le résultat de Yurinskii en affaiblissant les hypothèses de la façon suivante : nous supposerons toujours que l'inégalité (13) est vérifiée, mais nous remplacerons (15) par :

(17) $$V_r(P,Q) = \int_{\mathbb{R}^p} \|x\|^r |P-Q| dr < +\infty$$

alors

$$\Pi(P,Q) \leq C\left\{W^{B/\mathrm{Log}\,T}\,\frac{\mathrm{Log}\,T}{T} + \left(\int_{\|t\|<T} \sum_{(u)\leq r} \|D^u\Delta(t)\|^2 dt\right)^{\frac{1}{2}}\right\}$$

où B et C ne dépendent que de p , $T \geq e$ et $W = \sum_{j=0}^{r} V_j^2(P,Q)$.

Si l'on suppose que $p = 1$ et que P et Q vérifient (41), il existe C tel que pour $T > 0$

$$\Pi(P,Q) < C\left\{\frac{1}{T^{\frac{1}{2}}} + \left[\int_{|t|<T} (|\Delta(t)|^2 + |\Delta'(t)|^2) dt\right]^{\frac{1}{2}}\right\} .$$

Comme on le constate aisément sur l'inégalité (20) ces résultats nécessitent de nombreux calculs. Les auteurs n'ont pas précisé quelle était l'échelle de l'approximation obtenue.

18. - Distance de Prokhorov, distance de Lévy

Rappelons la définition de la distance de Lévy pour les probabilités sur $\mathbb{R}$. Soient P et Q deux probabilités sur $\mathbb{R}$ et F, G leurs fonctions de répartition respectives

$$L(P,Q) = \inf\{\epsilon / F(t) \leq G(t+\epsilon) + \epsilon\} .$$

En se restreignant à des probabilités à support contenu dans $[0,1]$, Bretagnolle et Huber [3] donnent le résultat suivant :

19. - PROPOSITION. - Soient P et Q deux probabilités sur $[0,1]$. Alors,

$\Pi^2 \leq L \leq \Pi$.

De plus, si l'une des deux probabilités admet une densité minorée, par rapport à la mesure de Lebesgue $\Pi = 0(L)$.

N.B. La propriété $L \leq \Pi$ est toujours vérifiée. Il peut y avoir égalité.

Pour terminer cette présentation, on rappelle ici deux résultats fondamentaux sur la distance de Prokhorov.

20. THEOREME (Prokhorov) : La convergence au sens de la distance de Prokhorov est équivalente à la convergence étroite des mesures.

Le second résultat connu sous le nom de théorème de Strassen a fait l'objet de plusieurs démonstrations (Strassen, Dudley, Schay).

Introduisons un espace probabilisé $(\Omega, \mathcal{F}, \gamma)$ et considérons l'ensemble $\mathcal{X}$ des variables aléatoires définies sur $(\Omega, \mathcal{F}, \gamma)$ à valeurs dans $(S, \mathcal{B})$. On peut munir $\mathcal{X}$ d'une distance "en probabilité" grâce à la distance de Ky Fan notée d_{pr} : si X et Y sont deux éléments de $\mathcal{X}$,

(21) $$d_{pr}(X,Y) = \inf\{\varepsilon / \gamma[d(X,Y) > \varepsilon] \leq \varepsilon\} .$$

Remarquons tout de suite la propriété suivante :

22. - LEMME. - Si X et Y sont deux variables aléatoires de lois respectives P et Q , alors $\Pi(P,Q) \leq d_{pr}(X,Y)$.

Démonstration. - Soit A un élément quelconque de $\mathcal{B}$. L'inclusion $[X^{-1}(A)] \cap [Y^{-1}(\overline{A^{\varepsilon}})] \subset [d(X,Y) > \varepsilon]$ est vérifiée. Alors

$$P(A) = \gamma[X^{-1}(A)] = \gamma[X^{-1}(A) \cap Y^{-1}(A^{\varepsilon})] + \gamma[X^{-1}(A) \cap Y^{-1}(\overline{A^{\varepsilon}})]$$
$$\leq \gamma[Y^{-1}(A^{\varepsilon})] + \gamma[d(X,Y) > \varepsilon]$$
$$\leq Q(A^{\varepsilon}) + \varepsilon .$$ On en conclut ce résultat.

On peut alors énoncer le théorème de Strassen :

23- THEOREME. - Soient (S,d) un espace métrique séparable et complet et P et Q deux lois de probabilités sur $(S,\mathcal{B})$. Alors il existe une probabilité λ sur $(S\times S, \mathcal{B}\otimes\mathcal{B})$ et des variables aléatoires X et Y de $(S\times S, \mathcal{B}\otimes\mathcal{B}, \lambda)$ dans $(S,\mathcal{B})$ de lois respectives P et Q telles que :

$\Pi(P,Q) = d_{pr}(X,Y)$.

24. - Calcul de la distance de Prokhorov. D'une manière générale, le calcul de la distance de Prokhorov entre deux probabilités quelconques P et Q n'est pas chose facile. Par contre, il existe de nombreux résultats d'encadrement de cette distance. Donnons tout de suite un premier résultat.

25. - LEMME. Soit A un borélien quelconque. On a les inégalités :

$$\epsilon(A) \leq \Pi(P,Q) \leq \sup_A |P(A) - Q(A)| .$$

Preuve. - L'inégalité de gauche résulte de la définition (9). Pour l'inégalité de droite, soit $\epsilon > 0$ tel que $\sup_A |P(A)-Q(A)| < \epsilon$. Alors $\forall A$, $P(A) - Q(A) \leq \sup_A |P(A)-Q(A)| < \epsilon$. Soit : $P(A) < Q(A) + \epsilon \leq Q(A^{\epsilon}) + \epsilon$. D'où le résultat.

La formule (9) nous incite à chercher si dans certains cas, il existe un borélien facilement identifiable A_o tel que $\Pi(P,Q) = \epsilon(A_o)$.

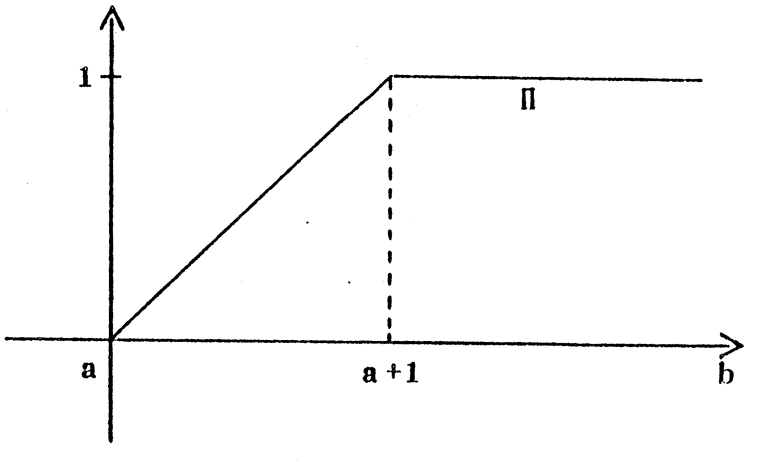
Le calcul de Π se ramène alors à la recherche de ce borélien "solution". Cependant ces cas semblent restreints. Rien n'empêche non plus de penser qu'en munissant $\mathcal{B}$ d'une distance appropriée (par exemple la distance de Hausdorff) et pour deux distributions P et Q suffisamment régulières (par exemple définies par des densités continues et de même support) $\epsilon(A)$ est une fonction continue de A. Ce qui devrait nous permettre d'obtenir un algorithme pour l'obtention de $\sup_A \epsilon(A)$. Avant de démontrer quelques résultats sur $(\mathbb{R},\mathcal{B})$, donnons deux exemples simples.

26. - <u>Exemples.</u> -

e_1) Si δ_a et δ_b sont deux masses de Dirac en a et b alors le borélien solution est $\{a\}$

et $\Pi(\delta_a, \delta_b) = \inf(1, d(a,b))$.

Si $S = \mathbb{R}$ et $d(a,b) = |b-a|$, on obtient la courbe suivante :



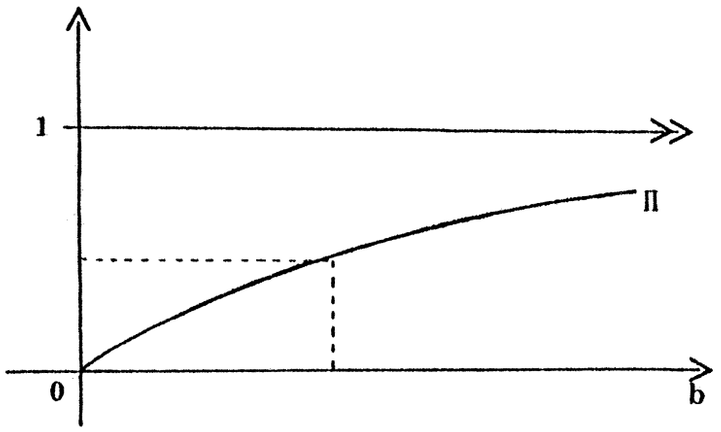
Distance de Prokhorov entre δ_a et δ_b en fonction de $b \in \mathbb{R}$.

On a donc $\Pi(\delta_o, \delta_{1/4}) = 1/4$. Si l'on effectue la transformation $h(x) = 2x$ on a $\Pi(\delta_{h(0)}, \delta_{h(1/4)}) = \Pi(\delta_o, \delta_{1/2}) = 1/2$: il n'y a pas invariance par transformation linéaire, car cette transformation n'est pas une isométrie.

e_2) Prenons encore $S = \mathbb{R}$, muni de la valeur absolue. Calculons la distance de Prokhorov entre la masse de Dirac δ_o et la loi uniforme G sur $[0,b]$. Le borélien solution est $\{0\}$. On doit donc avoir

$1 \leq G([0,\varepsilon]) + \varepsilon$, soit $1 \leq \frac{\varepsilon}{b} + \varepsilon = \varepsilon \frac{b+1}{b} \Leftrightarrow \varepsilon \geq \frac{b}{b+1}$

$\Pi(\delta_o, G) = \frac{b}{b+1}$.



Distance de Prokhorov entre δ_o et la loi uniforme sur $[0,b]$ en fonction de b .

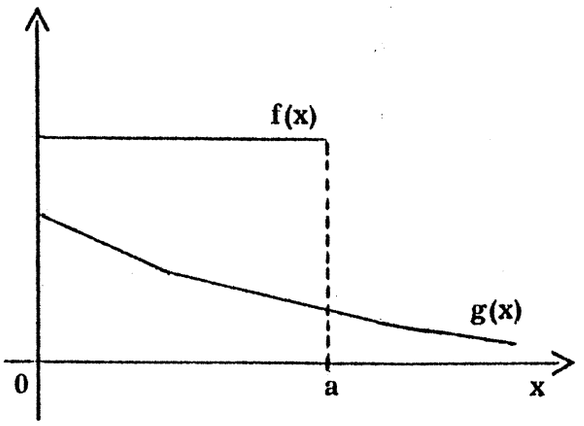
27. - PROPOSITION - Soient P et Q deux probabilités définies respectivement par les densités f et g . Soit $a > 0$. Nous supposerons que $A = [0,a]$ est le support de f . Nous supposons que g est décroissante sur $\mathbb{R}^+$ et qu'enfin sur A , $f > g$. Alors,

$\Pi(P,Q) = \varepsilon(A)$.

Preuve. - Commençons par illustrer de façon simple les hypothèses.

$f(x) = \frac{2}{3} \mathbb{1}_{[0,\, 3/2]}(x)$

$g(x) = \frac{1}{2} e^{-\frac{x}{2}} \mathbb{1}_{\mathbb{R}^+}(x)$



1) Posons $\epsilon(A) = \epsilon$ et montrons tout d'abord que si $B \subset A$, alors $\epsilon(B) \leq \epsilon$. On a en effet :

$$P(B) - Q(B) \leq P(A) - Q(A) \leq Q(A^{\epsilon} - A) + \epsilon .$$

Or, $A^{\epsilon} \cap \mathbb{R}^{+} = [0, a+\epsilon]$ et $Q(A^{\epsilon} - A) = Q(]a, a+\epsilon])$.

Soit $b = \sup B$. Alors, par hypothèse $b \leq a$ et donc

$$Q(B^{\epsilon} - B) \geq Q(]b, b+\epsilon]) \geq Q(]a, a+\epsilon])$$

d'après la décroissance de g. Donc

$$P(B) - Q(B) \leq Q(B^{\epsilon} - B) + \epsilon \Rightarrow \epsilon(B) \leq \epsilon .$$

2) Soit B un borélien quelconque de $\mathbb{R}^{+}$. Alors

$$P(B) = P(B \cap A) \leq Q[(B \cap A)^{\epsilon}] + \epsilon \leq Q(B^{\epsilon}) + \epsilon$$

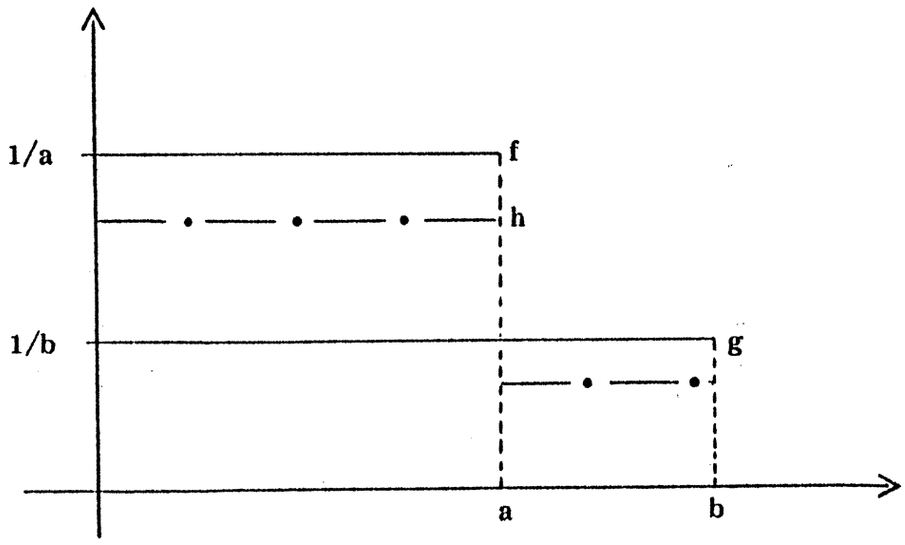
car $(B \cap A \subset B) \Rightarrow ((B \cap A)^{\epsilon} \subset B^{\epsilon})$

donc $\epsilon(B) \leq \epsilon$.

28. - <u>Exemples</u>. (cf. [15]).

e_3) Soient F, G, H les probabilités sur $\mathbb{R}$ définies par les densités respectives f, g, h où $f = \frac{1}{a} \mathbb{1}_{[0,a]}(x)$; $g = \frac{1}{b} \mathbb{1}_{[0,b]}(x)$ et $h = tf + (1-t)g$ avec $a > 0$, $b \geq a$ et $a < t \leq 1$. Calculons $\Pi(F, H)$: borélien solution : $A_o = [0, a]$. On a donc d'après la proposition 8 :

$$F[0,a] \leq H[0, a+\epsilon] + \epsilon \quad \text{soit} \quad 1 \leq t + \frac{(1-t)(a+\epsilon)}{b} + \epsilon$$



soit $\varepsilon \geq \dfrac{(b-a)(1-t)}{b+1-t}$ et $\Pi(F,H) = \dfrac{(b-a)(1-t)}{b+1-t}$.

En faisant $t = 0$, on obtient $\Pi(F,G) = b-a/b+1$; par des arguments similaires, $\Pi(G,H) = t(b-a)/b+1$.
(Si on fait $a = 0$ dans l'expression de $\Pi(F,G)$ on retrouve le résultat (e_2)).

e_4) Reprenons $f(x) = \frac{2}{3}\mathbb{1}_{[0,\,3/2]}(x)$ et $g(x) = \frac{1}{2}e^{-\frac{x}{2}}\mathbb{1}_{\mathbb{R}^+}(x)$.
$A = [0, \frac{3}{2}]$ et la condition sur ε s'écrit :

$$1 \leq 1 - e^{-(\frac{3}{4}+\frac{\varepsilon}{2})} + \varepsilon \ , \quad \text{soit} \quad \varepsilon - e^{-\frac{3}{4}} e^{-\frac{\varepsilon}{2}} \geq 0 \ .$$

On trouve :

$$\varepsilon(A) \ \# \ 0,389 \ .$$

On remarque que le borélien solution A vérifie :

$$P(A) - Q(A) = \sup_{B \in \mathcal{B}} |P(B)-Q(B)| \ ,$$

particularité qui permet, en général, de l'identifier facilement. On est tenté d'essayer de généraliser le résultat précédent à d'autres distributions de probabilités. En particulier à des probabilités P et Q définies sur $\mathbb{R}^+$ par des densités f et g toutes deux continues et décroissantes avec un seul point d'intersection a .

29. - LEMME - Avec les hypothèses précédentes pour P et Q, soit $b > a$, voisin de a. Une condition nécessaire pour que $\Pi(P,Q) = \varepsilon(A)$ est :

$$P(]a,b]) \leq Q([a+\varepsilon(A), b+\varepsilon(A)[) .$$

Preuve. - Par hypothèse $P(A) = Q(A^{\varepsilon(A)}) + \varepsilon(A)$ et $\forall B \in \mathcal{B}$, $P(B) \leq Q(B^{\varepsilon(A)}) + \varepsilon(A)$. Alors en prenant $B = [0,b]$ avec $a < b < a+\varepsilon(A)$ on doit avoir :

$$P(A) + P(]a,b]) \leq Q([0, a+\varepsilon(A)[) + Q([a+\varepsilon(A), b+\varepsilon(A)[) + \varepsilon(A) .$$

On en conclut le résultat.

Cette condition est vérifiée si g est constante à droite de a sur un intervalle suffisamment grand. Mais d'une façon générale, on constate qu'elle est rarement vérifiée.

30. - Exemple. - Soient P et Q les probabilités définies sur $\mathbb{R}^+$ par les densités respectives $f(x) = e^{-x}\mathbb{1}_{\mathbb{R}^+}(x)$ et $g(x) = \frac{1}{2}e^{-\frac{x}{2}}\mathbb{1}_{\mathbb{R}^+}(x)$. On trouve facilement que $a = 2\text{Log}\,2 \# 1,386$ et $\varepsilon(A) \# 0,202$. On a calculé en fonction de f les probabilités qui interviennent dans le lemme 29. (cf. tableau 30 bis). La condition n'est pas vérifiée.

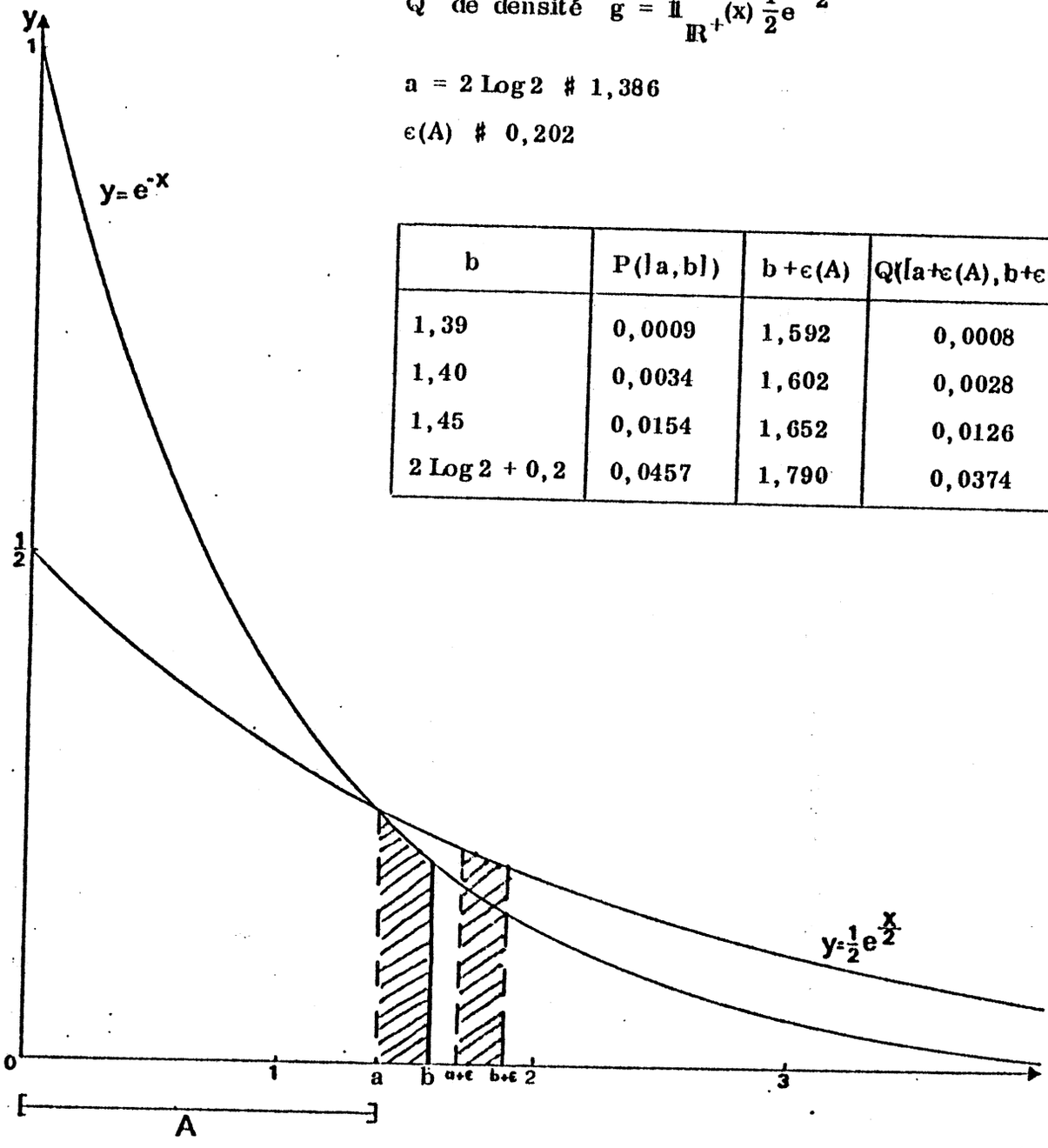
P de densité $f = \mathbb{1}_{\mathbb{R}^+}(x)e^{-x}$

Q de densité $g = \mathbb{1}_{\mathbb{R}^+}(x)\,\frac{1}{2}e^{-\frac{x}{2}}$

$a = 2\,\mathrm{Log}\,2 \;\#\; 1,386$

$\varepsilon(A) \;\#\; 0,202$

b	$P(]a,b])$	$b+\varepsilon(A)$	$Q([a+\varepsilon(A),b+\varepsilon(A)[)$
1,39	0,0009	1,592	0,0008
1,40	0,0034	1,602	0,0028
1,45	0,0154	1,652	0,0126
2 Log 2 + 0,2	0,0457	1,790	0,0374



- Graphe et tableau 30 bis -

31. - LEMME. - Soient P et Q deux probabilités portées par $\mathbb{R}^+$ et définies par les densités f et g toutes deux strictement décroissantes avec un seul point d'intersection a. Alors,

$$L(P,Q) = \Pi(P,Q)$$

où $L(P,Q)$ désigne la distance de Lévy entre P et Q.

Démonstration. - On sait que $L \leq \Pi$. Supposons donc $\Pi > L$. Soit $\epsilon > 0$ tel que $L < \epsilon < \Pi$. Alors $\exists B$ dont l'intersection avec $[0,a]$ est non vide tel que $P(B) > Q(B^\epsilon) + \epsilon$. L'ouvert B^ϵ est une union $\bigcup_j]a_j,b_j[$ d'intervalles ouverts disjoints de longueur $\geq 2\epsilon$. L'une au moins des bornes b_j peut être $+\infty$ et on adoptera la convention $+\infty - L = +\infty$. Alors $B \subset \bigcup_j]a_j+L, b_j-L[$ et l'on a :

$$\epsilon < P(B) - Q(B^\epsilon) < \sum_j P(]a_j+L, b_j-L[) - \sum_j Q(]a_j,b_j[) .$$

Soit

$$b_m = \begin{cases} b_{j_o} - L & \text{si } \exists j_o \text{ tel que } a \in]a_{j_o}+L, b_{j_o}-L[\\ \max\{b_j - L / b_j - L \leq a\} & \text{sinon.} \end{cases}$$

Alors

$$\epsilon < P[0,b_m] - Q[0,b_m+L] \leq L ,$$

ce qui conduit à une contradiction.

32 - LEMME. - Soient $a > 0$, $A = [0,a]$, P et Q deux probabilités sur $\mathbb{R}^+$ définies respectivement par les densités f et g et vérifiant :

1) $\forall x \in [0,a[\qquad f(x) > g(x)$;

2) $\forall x \in]a,+\infty[\qquad f(x) < g(x)$;

3) $\inf_{x\in[0,a[} g(x) \geq \sup_{x\in]a,+\infty[} g(x)$;

4) $\forall B \in \mathcal{B}(\mathbb{R}^+)$ tel que $B \cap A$ et $B \cap \bar{A}$ sont non vides, et tel que $P(B) - Q(B) > \epsilon(A)$, on a :

$$P(B\cap\bar{A}) \leq Q\left(B^{\varepsilon(A)} \cap \overline{(B\cap A)^{\varepsilon(B\cap A)}}\right) + \varepsilon(A) - \varepsilon(B\cap A) ,$$

<u>alors</u> $\Pi(P,Q) = \varepsilon(A)$.

<u>Preuve</u>. -

1) Si $B \subset \bar{A}$, alors $\varepsilon(B) = 0$. Si $B \subset A$, alors en utilisant le 1) de la preuve du lemme 14, compte-tenu de l'hypothèse 3), on a $\varepsilon(B) \leq \varepsilon(A)$.

2) Soit B un borélien de $\mathbb{R}^+$ tel que $B \cap A \neq \emptyset$. Si $P(B) - Q(B) \leq \varepsilon(A)$ a fortiori $P(B) \leq Q(B^{\varepsilon(A)}) + \varepsilon(A)$. Supposons donc que $P(B) - Q(B) > \varepsilon(A)$. Alors :

$$P(B) = P(B\cap A) + P(B\cap\bar{A}) \leq Q\left((B\cap A)^{\varepsilon(B\cap A)}\right) + \varepsilon(B\cap A) + P(B\cap\bar{A}) .$$

Utilisons la condition 4) : il vient

$$P(B) \leq Q\left((B\cap A)^{\varepsilon(B\cap A)}\right) + Q\left(B^{\varepsilon(A)} \cap \overline{(B\cap A)^{\varepsilon(B\cap A)}}\right) + \varepsilon(A)$$

mais

$$\left\{(B\cap A) \subset A\right\} \Rightarrow \left\{(B\cap A)^{\varepsilon(B\cap A)} \subset B^{\varepsilon(B\cap A)}\right\} ,$$

d'après 1) :

$$\varepsilon(B\cap A) \leq \varepsilon(A) ,$$

donc

$$B^{\varepsilon(B\cap A)} \subset B^{\varepsilon(A)} ,$$

d'où

$$B^{\varepsilon(A)} \cap (B\cap A)^{\varepsilon(B\cap A)} = (B\cap A)^{\varepsilon(B\cap A)} .$$

On en conclut :

$$P(B) \leq Q(B^{\varepsilon(A)}) + \varepsilon(A) ,$$

soit :

$$\varepsilon(B) \leq \varepsilon(A) .$$

Il est difficile de vérifier la condition 4). Elle concerne les boréliens qui coupent A et $\bar{A}$. Parmi de tels boréliens pour des densités f et g toutes deux décroissantes sur $\mathbb{R}^+$ et vérifiant 1) et 2), ceux de la forme $[0,b]$ jouent un rôle particulier. En effet, l'avance maximale de P sur Q est réalisée sur A . Q comble ensuite ce retard et pour b

suffisamment grand $P([0,b]) - Q([0,b]) < \varepsilon(A)$ et a fortiori $\varepsilon([0,b]) < \varepsilon(A)$. Donc, les boréliens de la forme $[0,b]$ avec b suffisamment proche de a sont ceux pour lesquels la condition 4) est le plus rarement vérifiée. Voyons alors ce que signifie cette condition.

Posons $[0,b] = B$. On a :

$$B \cap A = A \quad ; \quad B \cap \bar{A} =]a,b] \quad ; \quad \varepsilon(B \cap A) = \varepsilon(A)$$

$$(B \cap A)^{\varepsilon(A)} = [0, a+\varepsilon(A)[\quad ; \quad B^{\varepsilon(A)} \cap \overline{(B \cap A)^{\varepsilon(B \cap A)}} = [a+\varepsilon(A), b+\varepsilon(A)[\ .$$

La condition 4) s'écrit :

$$P(]a,b]) \leq Q([a+\varepsilon(A), b+\varepsilon(A)[)$$

qui est la condition nécessaire du lemme 16.

33. - <u>Exemple</u>. - Considérons la probabilité P définie par la densité $f(x) = e^{-x} \mathbb{1}_{\mathbb{R}^+}(x)$ et la probabilité Q définie par la densité

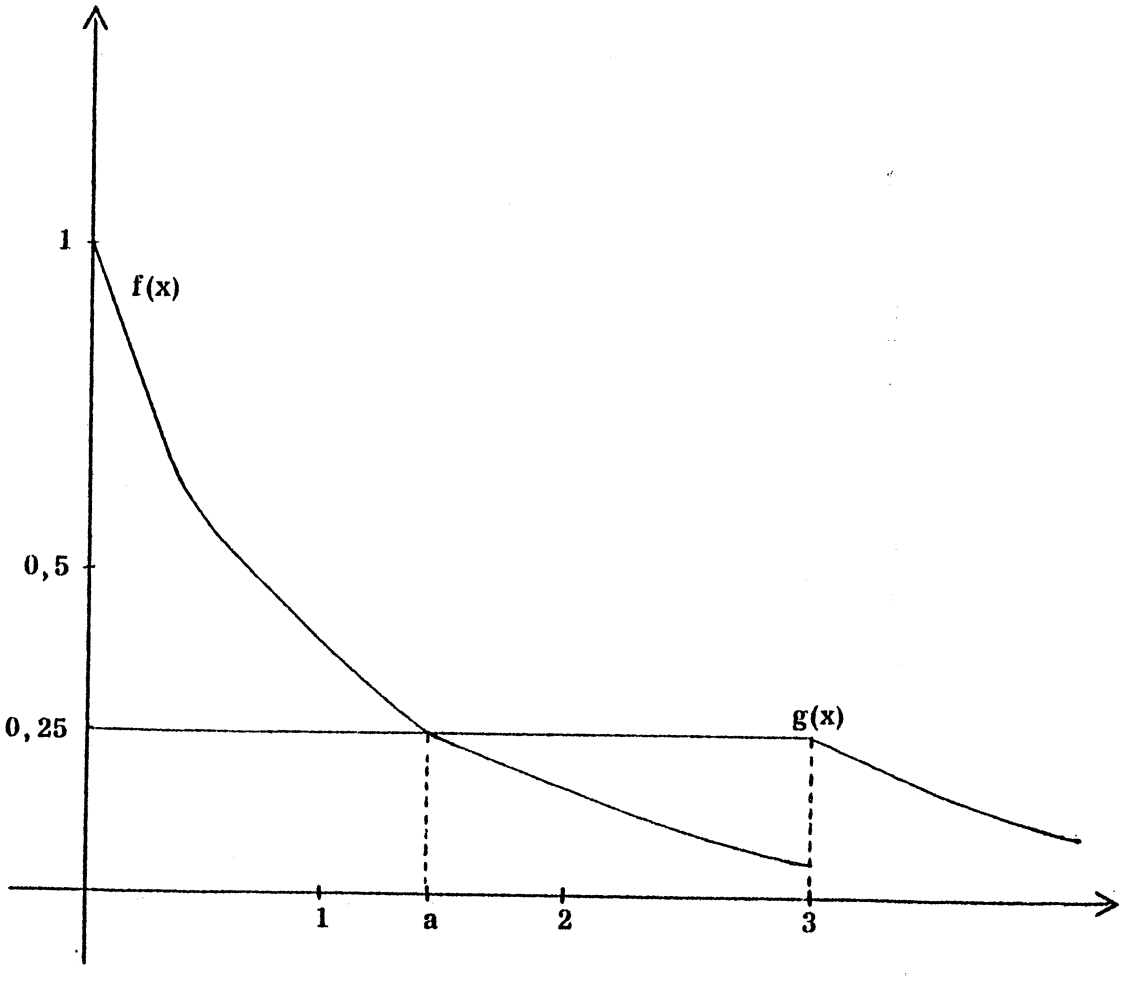
$$g(x) = \begin{cases} \frac{1}{4} & \text{si } 0 \leq x \leq 3 \\ e^{-(x-3+2\operatorname{Log}2)} & \text{si } x > 3 \ . \end{cases}$$

$$a = 2 \operatorname{Log} 2 \ , \qquad \varepsilon(A) = 0{,}32275^+$$

b	$P([a,b])$	$b+\varepsilon(A)$	$Q([a+\varepsilon(A), b+\varepsilon(A)[)$
1,4	0,00340	1,72275	0,00343
1,5	0,02697	1,82275	0,02843
1,6	0,0481	1,92275	0,05343
1,7	0,0673	2,02275	0,07843
1,8	0,0847	2,12275	0,10343

On vérifie alors que les conditions 1), 2), 3) et 4) du lemme 19 sont vérifiées et l'on a :

$$\Pi(P,Q) = \varepsilon(A) = 0{,}32275^+ \ .$$



Cependant, ces cas resteront limités. Et puisque dans un cas aussi simple que celui de l'exemple 31, on ne sait pas trouver le borélien solution, on peut se demander ce qu'il en est lorsque P et Q sont tout à fait quelconques.

34. - Calcul approché de la distance de Prokhorov.

On a alors l'idée suivante. Soient P et Q deux probabilités quelconques définies sur $\mathbb{R}$. On va construire deux suites de probabilités (P_n) et (Q_n) portées par des ensembles finis qui convergent faiblement respectivement vers P et Q . La convergence au sens de la distance de Prokhorov étant équivalente à la convergence faible des probabilités (cf. [13]), on sait que $\lim_{n\to+\infty} \Pi(P_n,Q_n) = \Pi(P,Q)$. On va alors estimer pour n suffisamment grand, fonction de la précision désirée, $\Pi(P,Q)$ par

$\Pi(P_n, Q_n)$. Mais ceci suppose d'une part que l'on puisse construire facilement les suites (P_n) et (Q_n) , d'autre part que l'on sache calculer $\Pi(P_n, Q_n)$ pour des ensembles de définition de l'ordre de 1 000 à 2 000 points. Enfin, il nous faut une majoration de l'erreur convenable. Nous allons voir comment l'on peut résoudre tous ces problèmes. Nous rappelons un lemme que Prokhorov (cf. [13]) a donné dans le cas d'un espace métrique séparable et complet.

35. - LEMME. - Soit P une probabilité définie sur $(\mathbb{R}, \mathcal{B})$. Soient φ une fonction de $\mathbb{R}$ dans $\mathbb{R}$ mesurable, $\delta > 0$ et $H = \{x / |\varphi(x) - x| \leq \delta\}$. Si $P(H) \geq 1 - \epsilon$ alors $\Pi(P, P_\varphi) \leq \max(\epsilon, \delta)$ où P_φ est la probabilité image de P par φ .

Preuve. - Vérifions que $A \cap H \subset \varphi^{-1}(A^{\delta]})$. Soit $x \in A \cap H$, alors $(|\varphi(x) - x| \leq \delta$ et $x \in A) \Rightarrow \varphi(x) \in A^{\delta]}$: soit $x \in \varphi^{-1}(A^{\delta]})$. Alors $P(A) = P(A \cap H) + P(A \cap \bar{H}) \leq P(A \cap \bar{H}) + P(\bar{H}) \leq P(\varphi^{-1}(A^{\delta]})) + \epsilon = P_\varphi(A^{\delta]}) + \epsilon$.

On en déduit le résultat à l'aide de la définition 2 et de la remarque 4.

Soit alors un entier $n > 0$ donné. Sur $\mathbb{R}$, on peut trouver un intervalle fermé borné K_n tel que $\max(P(\mathbb{R} - K_n), Q(\mathbb{R} - K_n)) < \frac{1}{2n}$. Soit k_{min} et k_{max} les extrémités de cet intervalle et $\ell = k_{max} - k_{min}$ sa longueur. On divise K_n en $[n\ell] = m + 1$ intervalles disjoints ; où $[n\ell]$ est la partie entière de $n\ell$. On pose alors :

$$I_j = \left[k_{min} + \frac{j-1}{n} , k_{min} + \frac{j}{n}\right[\qquad j = 1, \ldots, m$$

$$I_{m+1} = \left[k_{min} + \frac{m}{n} , k_{max}\right] .$$

Soit alors x_j le milieu de I_j , $j = 1, \ldots, m$ et $x_{m+1} = x_m + \frac{1}{n}$ (voir graphe 24). On construit l'application φ_n :

$$\forall x \in I_j \qquad \varphi_n(x) = x_j \qquad j = 1, \ldots, m+1$$

$$\forall x \in \mathbb{R} - K_n \qquad \varphi_n(x) = x_{m+1} .$$

Il est clair que φ_n est mesurable, et l'on a :

36. - COROLLAIRE. - Soit P_n (resp. Q_n) la mesure image de P (resp. Q) par φ_n. Alors $\Pi(P,P_n) \leq \frac{1}{2n}$ (resp. $\Pi(Q,Q_n) \leq \frac{1}{2n}$).

Calcul de $\Pi(P_n,Q_n)$.

On a vu (cf. exemple 13) que lorsque l'ensemble de définition de P ou de Q est réduit à un point, on sait calculer la distance de Prokhorov entre P et Q. Il en est de même si cet ensemble ne contient que quelques points. Mais les choses se gâtent au fur et à mesure que cet ensemble grossit pour devenir rapidement totalement impraticable. Nous allons donc utiliser des méthodes de calculs informatiques et pour cela il faut établir un lien entre le calcul de la distance de Prokhorov et un problème de recherche opérationnelle, ce qui nous permet alors d'utiliser des algorithmes connus.

Posons $S_n = \{x_1,\dots,x_{m+1}\}$. Pour $B \subset S_n$, on note $B_n^{\alpha]} = B^{\alpha]} \cap S_n$. On note (u,v) un élément de $\mathbb{R}^{2(m+1)}$ de coordonnées : u_i, $i=1,\dots,m+1$; v_i, $i=1,\dots,m+1$.

Soit $1 \geq \alpha > 0$ donné. A l'aide de la variante (10) de la définition, on voit que le calcul de la distance de Prokhorov entre P_n et Q_n nécessite la recherche de $\beta(\alpha) = \inf\{\beta / \forall B \in \mathcal{B},\ P_n(B) \leq Q_n(B^{\alpha]}) + \beta\}$.

37. - LEMME (cf. [18]). - Posons $P_n(x_i) = p_i$, $Q_n(x_i) = q_i$ et

$$d_{ij} = \begin{cases} 1 & \text{si } |x_i - x_j| \leq \alpha \\ 0 & \text{si } |x_i - x_j| > \alpha \end{cases} \qquad i,j = 1,\dots,m+1 .$$

(38) Alors sous les contraintes $u_i, v_j \geq 0$, $\forall i,j : u_i + v_j \geq d_{ij}$

(39) $$\beta(\alpha) = 1 - \min \sum_{i=1}^{m+1} (p_i u_i + q_i v_i) .$$

La minimisation de la forme linéaire sur $\mathbb{R}^{2(m+1)}$:

$$W(u,v) = \sum_{i=1}^{m+1} (p_i u_i + q_i v_i)$$

est un problème classique de programmation linéaire. Pour la démonstration du lemme 24, nous allons énoncer un résultat qui nous garantit l'existence d'une solution et la localise. Soient (p_i) , (q_i) , (d_{ij}) , $i,j=1,\dots,m+1$ des suites de réels ≥ 0 tels que $\sum_{i=1}^{m+1} p_i = \sum_{i=1}^{m+1} q_i$.

40. - LEMME (cf. [9]). - <u>Avec les notations précédentes, minimiser la forme linéaire</u> W <u>sur</u> $\mathbb{R}^{2(m+1)}$ <u>sous les contraintes (25) a une solution. Cette solution nous est donnée par la valeur de</u> W <u>sur l'un des sommets du parallélogramme</u> $\mathcal{V}$:

$$\mathcal{V} = \{(u,v)/0 \leq u_i \leq \max_j d_{ij} \text{ pour } i=1,\dots,m+1 \text{ et } 0 \leq v_j \leq \max_i d_{ij} \text{ pour } j=1,\dots,m+1\}$$

<u>situés dans la zone des contraintes.</u>

<u>Indication</u> : si l'on appelle C le polyèdre convexe déterminé par les contraintes (38), la démonstration consiste à prouver que $\inf\{W(u,v)/(u,v) \in C\} = \min\{W(u,v)/(u,v) \in \mathrm{Som}(C \cap \mathcal{V})\}$ où $\mathrm{Som}(C \cap \mathcal{V})$ désigne l'ensemble des sommets de $C \cap \mathcal{V}$, puis à prouver que $\mathrm{Som}(C \cap \mathcal{V}) \subseteq \mathrm{Som}(\mathcal{V})$. Pour plus de détails, nous renvoyons à [9].

<u>Démonstration du lemme 37</u>. - Compte-tenu des valeurs des d_{ij} le parallélogramme $\mathcal{V}$ défini au lemme 40 n'est autre que le cube unité de $\mathbb{R}^{2(m+1)}$. Un sommet de ce cube est caractérisé par un couple (u,v) dont les coordonnées valent 0 ou 1 . A toute suite $(u_i)_{i=1,\dots,m+1}$ de 0 ou 1 est associé de façon unique un ensemble $B \subset S_n$ tel que $u_i = 0 \Leftrightarrow x_i \in B$ et $u_i = 1 \Leftrightarrow x_i \in S_n - B$. Dans ces conditions

$$d_{ij} - u_i = \begin{cases} 1 & \text{si} \quad x_i \in \{x_j\}_n^{\alpha]} \cap B \\ 0 & \text{si} \begin{cases} x_i \in \{x_j\}_n^{\alpha]} \cap (S_n - B) \\ \text{ou} \\ x_i \in S_n - \{x_j\}_n^{\alpha]} \cap B \end{cases} \\ -1 & \text{si} \quad x_i \in S_n - \{x_j\}_n^{\alpha]} \cap (S_n - B) \ . \end{cases}$$

Les valeurs v_i qui respectent les contraintes sont donc données par :

$$v_i \geq \begin{cases} 1 & \text{si} \quad x_i \in B_n^{\alpha]} \\ 0 & \text{si} \quad x_i \in S_n - B_n^{\alpha]} \ . \end{cases}$$

En conclusion, la suite $\forall i$,

$$u_i = \begin{cases} 0 & \text{si} \quad x_i \in B \\ 1 & \text{si} \quad x_i \in S_n - B \end{cases} \qquad v_i = \begin{cases} 1 & \text{si} \quad x_i \in B_n^{\alpha]} \\ 0 & \text{si} \quad x_i \in (S_n - B_n^{\alpha]}) \end{cases}$$

caractérise un sommet x_B du cube unité qui est situé dans la zone des contraintes. Pour une telle suite :

$$\sum_i (p_i u_i + q_i v_i) = P_n(S_n - B) + Q_n(B_n^{\alpha]}) \ .$$

D'autre part, si x est un autre sommet du cube unité vérifiant les contraintes et tel que $u_i = 1$ pour tout $x_i \in S_n - B$, alors

$$W(x) \geq W(x_B) \ ,$$

car on a choisi pour caractériser x_B les plus petites valeurs des v_j qui vérifient les contraintes. Donc, d'après le lemme 40

$$\min W(u,v) = \min_{B \subset S_n} [P_n(S_n - B) + Q_n(B_n^{\alpha]})] \ .$$

Alors,

$$\beta(\alpha) = 1 - \min W(u,v) = 1 - \min_{B \subset S_n} [P_n(S_n - B) + Q_n(B_n^{\alpha]})]$$

$$= \max_{B \subset S_n} [P_n(B) - Q_n(B_n^{\alpha]})] = \inf\{\beta / \forall B \subset S_n \ , \ P_n(B) \leq Q_n(B_n^{\alpha]}) + \beta\} \ .$$

Cette dernière égalité étant immédiate.

Enfin, si B est un borélien quelconque on voit que

$$P_n(B) = P_n(B \cap S_n) \qquad \text{et} \qquad Q_n(B^{\alpha]}) = Q_n(B^{\alpha]} \cap S_n)$$

d'où le résultat.

Le problème se ramène donc à minimiser une forme linéaire. Il se trouve que ceci peut se résoudre sur ordinateur grâce à un algorithme de maximisation de flot : l'algorithme de Ford et Fulkerson (cf. [8]). Etant donné un sommet (x_i, p_i) on est capable grâce aux contraintes (38) de connaître les successeurs (x_j, q_j) de ce sommet. Grâce à un marquage direct et inverse l'algorithme de Ford et Fulkerson trouve ce flot noté $F\ell$ max . Cet algorithme nous libère des outils classiques de la programmation linéaire (algorithme du simplexe) et entraîne un gain en temps calcul et en places mémoires considérable.

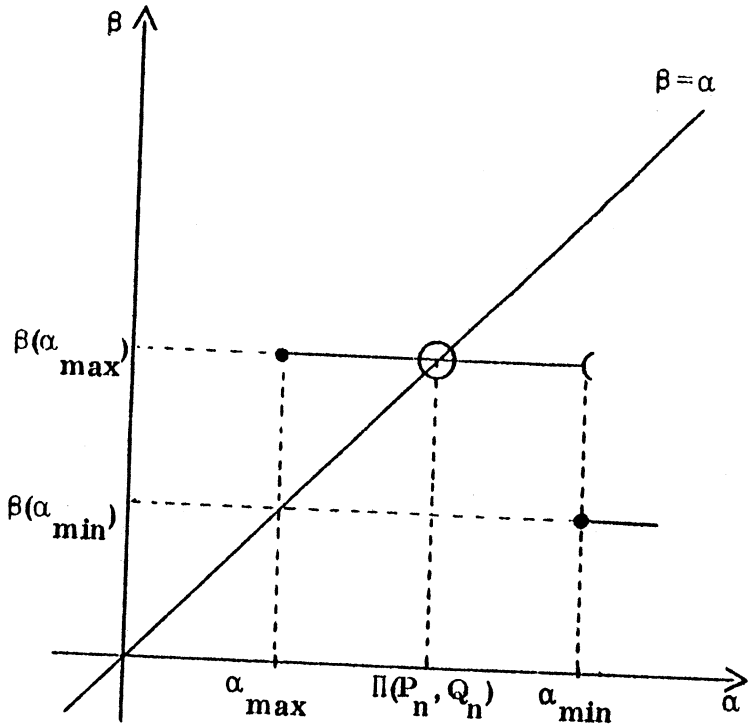
Compte-tenu de la construction de Q_n l'application $\alpha \to Q_n(B_n^{\alpha]})$ a des sauts pour des valeurs α de la forme k/n , $k=1,\dots,m$. Mais pour des gains de temps calcul, on cherche à restreindre le plus possible l'éventail des valeurs α pour lesquelles on calculera $\beta(\alpha)$. On sait que $0 \leq \Pi(P_n, Q_n) \leq 1$. Mais le lemme 25 nous permet en général un encadrement nettement meilleur. Pour des valeurs convenables de α on trouve donc $\beta(\alpha) = 1 - F\ell\max(\alpha)$ puis :

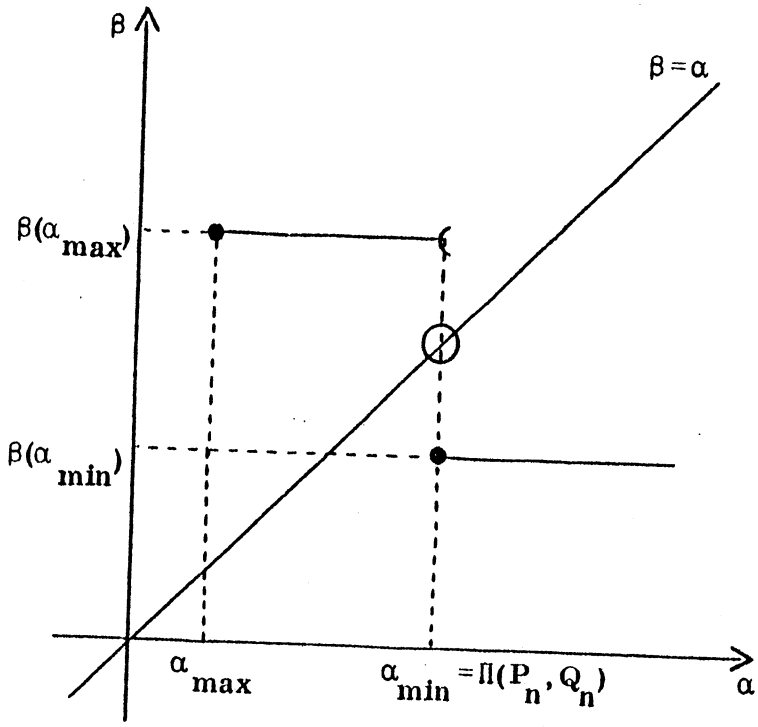
$$\alpha_{min} = \min\{\alpha/\alpha \text{ de la forme } k/n \text{ tel que } \beta(\alpha) \leq \alpha\}$$
$$\alpha_{max} = \max\{\alpha/\alpha \text{ de la forme } k/n \text{ tel que } \beta(\alpha) > \alpha\} .$$

En remarquant que $\beta(\alpha)$ est une fonction décroissante de α on trouve que :

(41) $$\inf\{\alpha/\beta(\alpha) \leq \alpha\} = \min[\alpha_{min}, \beta(\alpha_{max})] .$$

Les graphes (42) montrent les deux cas possibles.


β
β = α
β(α max)
β(α min)
α max
Π(P n, Q n)
α min
α


β
β = α
β(α max)
β(α min)
α max
α min = Π(P n, Q n)
α

- Graphes 42 -

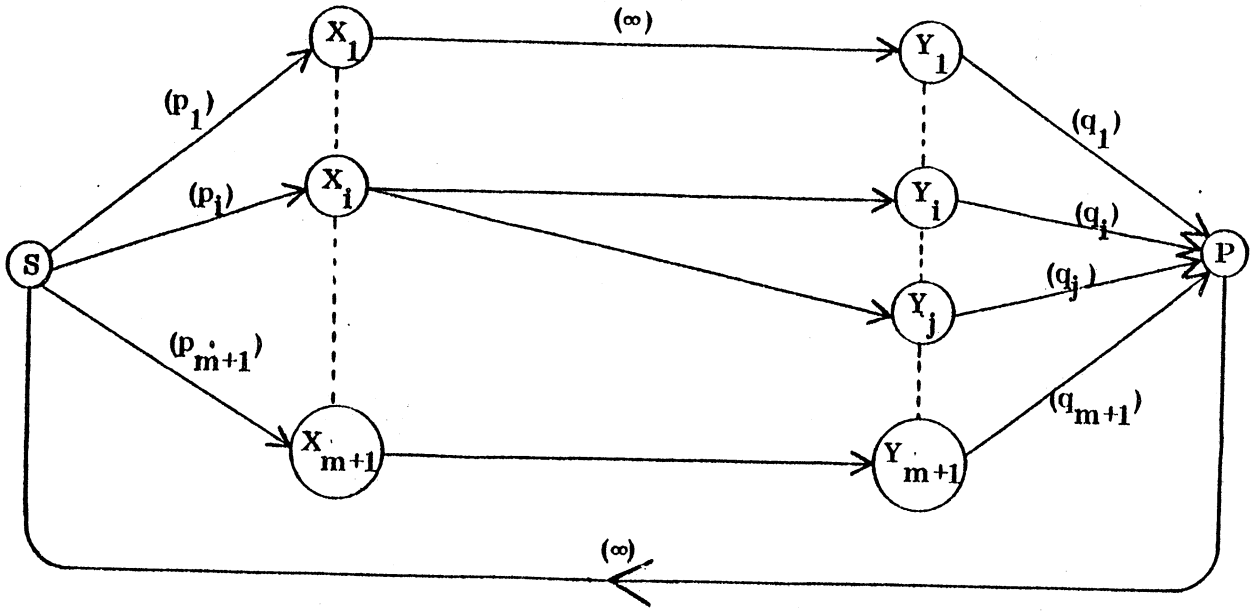
On peut enfin, grâce à l'inégalité triangulaire, trouver une majoration de l'erreur d'approximation :

$$\Pi(P_n,Q_n) - \Pi(P_n,P) - \Pi(Q_n,Q) \leq \Pi(P,Q) \leq \Pi(P,P_n) + \Pi(P_n,Q_n) + \Pi(Q_n,Q) .$$

(43) soit :

$$|\Pi(P,Q) - \Pi(P_n,Q_n)| \leq 1/n .$$

44.- <u>Exemple</u>. - Reprenons comme première illustration de cette méthode, les probabilités P et Q de l'exemple 17. Prenons $n = 100$. On trouve alors $K_n = [0 ; 10,597]$, $m+1 = 1060$, $I_{m+1} = [10,59 ; 10,597]$. Les autres intervalles ont une longueur de $\frac{1}{100}$. Les probabilités P_n et Q_n sont définies sur un ensemble de 1060 points ; chaque point étant équidistant des deux points qui l'entourent à une distance de $\frac{1}{100}$. Le calcul de $\int_{(i-1)/n}^{i/n} e^{-x}dx$ pour $i=1,\ldots,m$ est effectué en double précision, puis multiplié par 2^{34}, produit dont on ne garde que la partie entière et que l'on pose $= p_i$, de façon à n'avoir pour p_i que des valeurs entières. On procède de la même manière pour le calcul des q_i. La somme des erreurs d'arrondis est dans les deux cas inférieure à 4×10^{-8} et reportée sur p_{m+1}. Ceci n'affecte pas la précision du résultat. Le calcul de $\beta(\alpha)$ est effectué en quelques secondes par l'ordinateur. Pour expliquer les modalités de ce calcul, construisons le graphe suivant :



Sur les arcs de ce graphe, nous faisons circuler des flots bornés sur chaque arc par le nombre indiqué entre parenthèses. Montrons que le flot maximum entre S et P prend pour valeur $\min \sum_i (p_i u_i + q_i v_i)$ sous les contraintes $u_i \geq 0$, $v_i \geq 0$, $u_i + v_j \geq d_{ij}$ où $d_{ij} \geq 0$. En effet, le dual de ce problème est :

$$(D) \begin{cases} \text{maximiser} \quad \sum_{i,j} d_{ij} x_{ij} \\ \text{sous les contraintes :} \quad \sum_i d_{ij} x_{ij} \leq q_j \quad ; \quad \sum_j d_{ij} x_{ij} \leq p_i \\ x_{ij} \geq 0 \ . \end{cases}$$

Nous avons :

$$d_{ij} = 1 \quad \text{si} \quad |i-j| \leq n\alpha$$
$$d_{ij} = 0 \quad \text{si} \quad |i-j| > n\alpha \ .$$

Si $d_{ij} = 0$, on peut supprimer la variable x_{ij} correspondante. D'où (D) s'écrit :

$$\begin{cases} \text{maximiser} \quad \sum_{\substack{i,j \\ |i-j| \leq n\alpha}} x_{ij} \quad \text{où} \quad x_{ij} \geq 0 \\ \text{et} \quad \sum_{\substack{i \\ |i-j| \leq n\alpha}} x_{ij} \leq q_j \ , \quad \sum_{\substack{j \\ |i-j| \leq n\alpha}} x_{ij} \leq p_i \end{cases}$$

problème que décrit le graphe ci-dessus si x_{ij} est la valeur du flot allant de X_i à Y_j .

La capacité maximum de l'arc (S, X_i) exprime que $\sum_{\substack{i \\ |i-j| \leq n\alpha}} x_{ij} \leq p_i$. La capacité maximum de l'arc (Y_j, P) exprime que $\sum_{\substack{i \\ |i-j| \leq n\alpha}} x_{ij} \leq q_j$. Pour

Pour trouver le flot maximum, on utilise l'algorithme de Ford et Fulkerson (cf. [3]) qui est suffisamment connu pour que nous n'en fassions qu'un descriptif sommaire dans le cas qui nous intéresse. On appelle sommet insaturé : un sommet X_i tel que la somme des flots sortants est strictement inférieure à p_i , ou un sommet Y_j tel que la somme des flots entrants

est strictement inférieure à q_j . A chaque itération, on cherche une suite d'arcs partant d'un sommet insaturé X pour aller à un sommet insaturé Y . Cette suite d'arcs est alternée ; les arcs parcourus dans le sens contraire à l'orientation (du côté Y vers le côté X) doivent porter un flot non nul.

Si une telle suite existe, on peut augmenter le flot total. La valeur de l'augmentation est entière puisqu'obtenue par somme ou soustraction d'entiers. Ceci permet de borner aisément le nombre d'itérations, car le flot maximum est borné par $\sum_i p_i$ soit 2^{34} . Le nombre d'itérations est en réalité bien inférieur.

Au lieu de partir d'un flot nul, il est avantageux de partir d'un flot aussi grand que possible et ce flot nous est fourni par $\sum_{i=1}^{m+1} \min(p_i, q_i)$. On relie cette valeur à l'inégalité (8) de la façon suivante : d'après le lemme 6 et l'inégalité (8)

$$F\ell_{max}(\alpha) = 1 - \beta(\alpha) \geq 1 - \sup_B |P_n(B) - Q_n(B)|$$

$$= 1 - \sum_{J_1} (p_i - q_i) \quad \text{où} \quad J_1 = \{i / p_i > q_i\} \ , \quad J_2 = \{i / p_i \leq q_i\}$$

$$= (1 - \sum_{J_1} p_i) + \sum_{J_1} q_i = \sum_{J_2} p_i + \sum_{J_1} q_i$$

$$= \sum_i \min(p_i, q_i) \ .$$

Le tableau ci-dessous résume les caractéristiques de notre exemple et les résultats du programme.

$\sup_A |P(A) - Q(A)| = 0,25$. Pour $x_o = 1,590182$, $\epsilon([0, x_o]) = 0,203888$

$n = 100$, $m+1 = 1060$, $K_n = [0 ; 10,597]$.

	α	$F\ell_{max}(\alpha)$	$\beta(\alpha)$	$\Pi(P_n, Q_n)$
	.19	0,79326	0,20614	
$\alpha_{max} \rightarrow$	.20	0,79532	0,20468	
$\alpha_{min} \rightarrow$	.21	0,79735	0,20265	0,20468
	.22	0,79937	0,20063	

45. - <u>Prenons un deuxième exemple</u>. Dans ce qui suit P et Q sont des lois $N(0,\sigma)$ avec $\sigma \geq 1$. Notons tout d'abord qu'il serait vain d'espérer utiliser la symétrie des densités pour réduire l'étude à $\mathbb{R}^+$. Appelons en effet, P^+ et Q^+ la restriction de telles lois à $\mathbb{R}^+$ et P^- et Q^- leurs restrictions à $\mathbb{R}^-$. Soient $a > 0$ et ϵ_1 tel que

$$P^+([0,a]) \leq Q^+([0, a+\epsilon_1)[) + \epsilon_1$$

donc :

$$P^-([-a,0]) \leq Q^-(]-a-\epsilon_1,0]) + \epsilon_1$$

alors,

$$P([-a,+a]) = P^+([0,a]) + P^-([-a,0]) \leq Q^+[0,a+\epsilon_1] + Q^-[-a-\epsilon_1,0] + 2\epsilon_1$$
$$= Q([-a,+a]^{\epsilon_1} + 2\epsilon_1]$$

Soient P et Q deux lois normales centrées de variances respectives σ_1^2 , σ_2^2 avec $\sigma_1 > \sigma_2$. Appelons t_o le point d'intersection d'abscisse > 0 des graphes des densités. Il est facile de voir que

$$\sup_{A \in \mathcal{B}} |P(A) - Q(A)| = P([-t_o,t_o[) - Q([-t_o,t_o[)$$

et que t_o nous est fourni par :

$$\frac{1}{\sqrt{2\Pi}\sigma_1} \exp\left\{-\frac{1}{2}\frac{t^2}{\sigma_1^2}\right\} = \frac{1}{\sqrt{2\Pi}\sigma_2} \exp\left\{-\frac{1}{2}\frac{t^2}{\sigma_2^2}\right\} \Leftrightarrow \frac{\sigma_2}{\sigma_1} = \exp\left\{\frac{t^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\right\}$$

soit $$t_o = \sqrt{2\,\frac{\operatorname{Log}\frac{\sigma_2}{\sigma_1}}{1/\sigma_1^2 - 1/\sigma_2^2}} \quad .$$

Si l'on appelle Φ la fonction $\int_0^t \frac{1}{\sqrt{2\Pi}} e^{-x^2/2} dx$ qui tabule la loi normale on a :

$$P([-t_o,t_o[) - Q([-t_o,t_o[) = 2\left[\Phi\left(\frac{t_o}{\sigma_1}\right) - \Phi\left(\frac{t_o}{\sigma_2}\right)\right] \quad .$$

Si l'on cherche $\epsilon([-t_o,t_o[)$, on a à résoudre $2\Phi\left(\frac{t_o}{\sigma_1}\right) - 2\Phi\left(\frac{t_o+\epsilon}{\sigma_2}\right) = \epsilon$ soit une équation de la forme $\Psi(\epsilon) = \epsilon$ avec $\Psi'(\epsilon) = \frac{1}{\sigma_2} e^{-\epsilon^2}$ qui justifie l'obtention d'une solution par une méthode itérative.

Le tableau 46 nous donne les valeurs numériques des inégalités (8).

Nous avons alors alternativement choisi $n = 10$ et $n = 50$. L'intervalle k_n est symétrique par rapport à l'origine et nous donnons au tableau la valeur k_{max}. On a alors divisé k_n en $2([n.k_{max}]+1)$ intervalles. On a reporté $\int_{-\infty}^{k_{min}} f(x)dx$ sur p_o et $\int_{k_{max}}^{+\infty} f(x)dx$ sur P_{2m+2}. Ces valeurs ainsi que $\int_{(i-1)/n}^{i/n} f(x)dx$ sont calculées par l'ordinateur au moyen de la formule d'approximation polynomiale (cf. [1]). Le programme a tourné avec des valeurs entières comme dans le cas II-1). Les résultats sont consignés au tableau 46 bis et nous avons pu tracer la courbe 47.

CALCULS PREALABLES AUX CALCULS SUR ORDINATEUR

P : N(0,1) Q : N(0,σ) $f(x) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{x^2}{2}}$ $g(x) = \frac{1}{\sqrt{2\Pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$

t_o point d'intersection >0 de f et g

Nous posons $A_o = [-t_o, +t_o]$, $\varepsilon(A_o) = \varepsilon_o$, $t_o = [2 \operatorname{Log}\sigma/(1-\frac{1}{\sigma^2})]^{1/2}$

Tableau 46

σ	t_o	$\operatorname{Sup}\lvert P(A)-Q(A)\rvert = P(A_o) - Q(A_o)$	$P(A_o)$	$\varepsilon_o + Q(A_o^{\varepsilon_o})$	ε_o
1,2	1,0924	0,0880	0,7253	0,7260	0,062
2	1,3596	0,3226	0,8260	0,8265	0,248
3	1,5722	0,48432	0,8841	0,8854	0,396
4	1,7197	0,58177	0,9145	0,9152	0,495
5	1,8311	0,64712	0,9329	0,9332	0,565
6	1,9199	0,69411	0,9451	0,9457	0,618
7	1,9932	0,72960	0,9538	0,9542	0,659
8	2,0555	0,75739	0,9602	0,9607	0,691
9	2,1094	0,77978	0,9651	0,9657	0,719
10	2,1598	0,79413	0,9690	0,9690	0,741

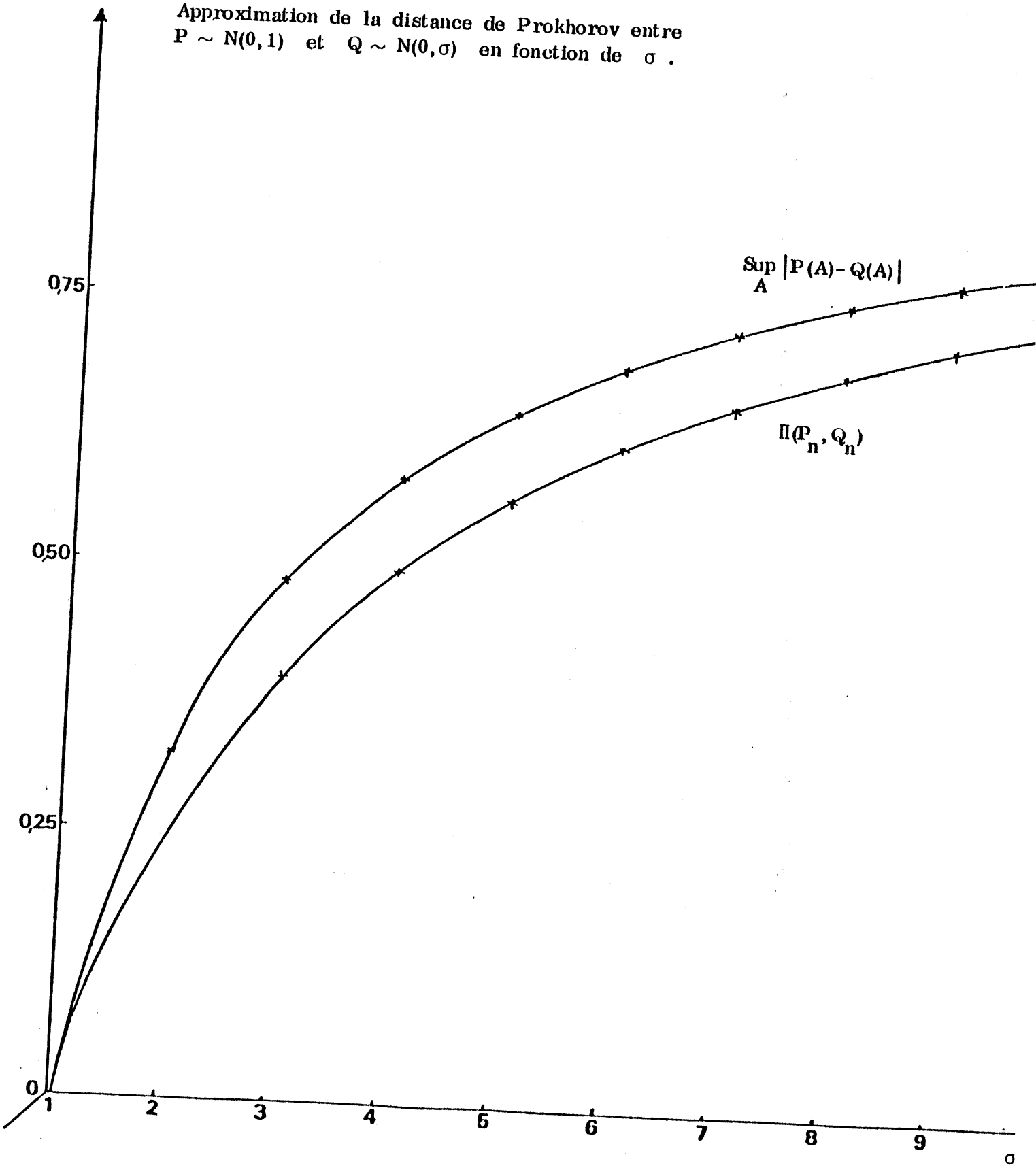
RESULTATS DES CALCULS SUR ORDINATEUR

Tableau 46bis

α	n	abscisse max.	α	$F\ell_{max}$	$\beta = 1 - F\ell_{max}$	$\Pi(P_n, Q_n)$
3	10	5,88	.39	0,58314	0,41686	
"	"	"	.40	0,60468	0 39532	.4
"	50	7,73	.39	0,60024	0,39976	
"	"	"	.40	0,60448	0,39976	.39976
4	10	7,84	.49	0,48921	0,51079	
"	"	"	.50	0,50636	0,49364	.5
"	50	10,30	.49	0,50293	0,49707	
"	"	"	.50	0,50633	0,49367	.49707
5	10	9,80	.56	0,42621	0,57379	
"	50	12,88	.56	0,43450	0,56550	
"	"	"	.58	0,43794	0,56266	.56550
6	50	15,46	.60	0,38033	0,61967	.61967
7	10	13,72	.6	0,33518	0,66482	.66482
"	50	18,03	.64	0,33913	0,66057	
"	"	"	.66	0,34155	0,65845	.66
8	50	20,61	.69	0,30742	0,69258	.69258
9	10	17,64	.7	0,28007	0,71993	.71993
10	10	19,60	.7	0,25599	0,74400	.74400

Courbe 47

Approximation de la distance de Prokhorov entre $P \sim N(0,1)$ et $Q \sim N(0,\sigma)$ en fonction de σ .



Remarque. - Pour $n = 10$, $\sigma = 3$ le flot maximum est obtenu en 26 itérations et le calcul a nécessité 1,94 secondes.

Pour $n = 50$, $\sigma = 3$ le flot maximum est obtenu en 125 itérations et le calcul a nécessité 29,94 secondes.

Remarque. - La mise en oeuvre de l'algorithme de Ford et Fulkerson n'impose nullement que les points $x_1, \ldots, x_{m+1}$ chargés par P_n soient les mêmes que ceux chargés par Q_n , ni que les nombres de ces points soient identiques. Le descriptif précédent qui convient particulièrement lorsque P et Q ont des densités de même support, pourra être modifié lorsque les supports sont disjoints ou que l'une au moins des deux lois est discrète. En particulier, lorsque P et Q sont des lois sur des ensembles finis quelconques, l'algorithme de Ford et Fulkerson permet de calculer la valeur exacte de leur distance de Prokhorov.

Cette méthode s'étend facilement à $S = \mathbb{R}^p$ avec $p > 1$ et plus généralement à S un espace métrique séparable et complet pour lequel on peut déterminer facilement pour tout x , l'ensemble $\{y \in S / d(x,y) \leq \alpha\}$, puisque l'algorithme ne nécessite que la connaissance des p_i et des capacités d'absorbtion des p_i (capacités des successeurs).

Quelques résultats complémentaires.

Sur $\mathbb{R}$, on prendra $d(x,y) = |x-y|$, sur $\mathbb{R}^p$, $d(x,y) = \max_{i=1,\ldots,p} |x_i - y_i|$.

Soient $x_1, \ldots, x_n$ (resp. $y_1, \ldots, y_n$) un n-échantillon d'une loi de probabilité P (resp. Q) . Soit P_n (resp. Q_n) la loi empirique associée.

48. - LEMME. - Supposons les x_i tous distincts et qu'il existe k indices $i_1, \ldots, i_k$ avec $1 \leq k \leq n$ tels que $|x_{i_j} - y_{i_j}| < \varepsilon$ pour $j = 1, \ldots, k$. Alors $\Pi(P_n, Q_n) \leq \max(\varepsilon, \lambda)$ où $\lambda = \frac{n-k}{n}$.

Preuve. - Soit φ l'application de $\mathbb{R}$ dans $\mathbb{R}$ telle que $\varphi(x_i) = y_i$ pour $i = 1,\dots,n$. Alors, Q_n est la probabilité image de P_n par φ et comme $P_n\{x/|\varphi(x)-x|<\epsilon\} \geq 1-\lambda$, on obtient le résultat à l'aide du lemme 22.

Remarque. - Ce résultat reste vrai si les x_i ne sont pas tous distincts, mais la démonstration précédente ne s'applique plus. Il faut alors revenir à la définition. Soit $B \in \mathcal{B}$. On peut supposer que $B \cap \{x_1,\dots,x_n\} \neq 0$ et $B \cap \{x_{i_1},\dots,x_{i_k}\} = \{x_{i_1},\dots,x_{i_\ell}\}$, alors $B^\epsilon \supset \{y_{i_1},\dots,y_{i_\ell}\}$. Donc :

$$P_n(B) \leq \frac{\ell}{n} + \lambda \leq Q_n(B^\epsilon) + \lambda$$

d'où le résultat.

49. - COROLLAIRE. - Supposons que $\max_{i=1,\dots,n} |x_i-y_i| < \epsilon$. Alors $\Pi(P_n,Q_n) \leq \epsilon$.

(Preuve. - Prendre $k = n$ et $\lambda = 0$ dans le lemme précédent).

Appelons maintenant $\mathcal{G}$ le groupe des permutations de coordonnées de $\mathbb{R}^n$ et $\tilde{\mathbb{R}}^n$ l'espace quotient. Soit H l'application canonique de $\mathbb{R}^n$ sur $\tilde{\mathbb{R}}^n$. La distance (25) est invariante pour $\mathcal{G}$. Il est possible de définir une distance $\tilde{d}$ sur $\tilde{\mathbb{R}}^n$ par

$$d(x,y) = \min_{g,h} d(g(x),h(y)) = \min_g d(x,g(y)) .$$

On obtient en effet l'inégalité triangulaire de la façon suivante. Soient x,y,z trois points de $\mathbb{R}^n$ et g,h deux éléments de $\mathcal{G}$. On a

$$d(g(x),h(y)) \leq d(g(x),z) + d(z,h(y))$$

d'où $\min_g d(g(x),h(y)) \leq \min_g d(g(x),z) + d(z,h(y))$

et $\min_h \min_g d(g(x),h(y)) \leq \min_g d(g(x),z) + \min_h d(z,h(y))$

soit : $\tilde{d}(\tilde{x},\tilde{y}) \leq \tilde{d}(\tilde{x},\tilde{z}) + \tilde{d}(\tilde{z},\tilde{y})$.

La réflexivité et la symétrie sont évidentes. Vérifions alors que $[H^{-1}(B)]^{\epsilon} \subset H^{-1}(B^{\epsilon})$ pour B borélien de $\hat{\mathbb{R}}^h$. Soit $x \in [H^{-1}(B)]^{\epsilon}$. Alors $\exists y \in H^{-1}(B)$ tel que $d(x,y) < \epsilon$. $y \in H^{-1}(B) \Leftrightarrow H(y) \in B$. De plus, $\tilde{d}(\tilde{x},\tilde{y}) = \inf_{g,h} d(g(x),h(y)) \leq d(x,y)$ obtenue pour $g = h =$ identité. Donc :

$$\tilde{d}(\tilde{x},\tilde{y}) < \epsilon ,$$

d'où :

$$\tilde{x} = H(x) \in B^{\epsilon} .$$

Soit P une probabilité sur $[\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)]$ et $\tilde{P}$ sa symétrisée.

50. - LEMME. - $\Pi(\tilde{P},\tilde{Q}) \leq \Pi(P,Q)$.

Preuve. - Soit $\epsilon > \Pi(P,Q)$ et soit $B \in \mathcal{B}(\mathbb{R}^n)$, alors

$$\tilde{P}(B) = P(H^{-1}(B)) < Q([H^{-1}(B)]^{\epsilon}) + \epsilon$$
$$\leq Q(H^{-1}(B^{\epsilon})) + \epsilon = \tilde{Q}(B^{\epsilon}) + \epsilon ,$$

d'où le résultat.

51. - LEMME. - Soient P et Q deux lois de probabilités symétriques par rapport à 0. Alors pour tout borélien B, il existe un borélien symétrique B_S tel que $\epsilon(B_S) \geq \epsilon(B)$.

On veut montrer que

$$\inf \{ \epsilon \ / \ P(B) \leq Q(B^{\epsilon}) + \epsilon \}$$

$$\leq \inf \{ \epsilon \ / \ P(B_S) \leq Q(B_S{}^{\epsilon}) + \epsilon \}$$

Soit $\epsilon > 0$. Posons pour tout borélien B : $B^+ = B \cap \mathbb{R}^+_*$
de même $B^- = B \cap \mathbb{R}^-_*$

On remarque que : $[(B_+)^{\epsilon}]_+ \subset (B^{\epsilon})_+$ est toujours vérifié et qu'il existe des cas où l'inclusion est stricte. Alors on peut vérifier que si l'on appelle B_S le borélien symétrique engendré par B_+ on a :

$(B_S{}^{\epsilon})_+ \subset (B^{\epsilon})_+$. On peut écrire :

$$P(B) - Q(B^{\varepsilon}) = P(B_+) - Q[(B^{\varepsilon})_+] + P(B_-) - Q[(B^{\varepsilon})_-]$$

Supposons que $P(B_+) - Q[(B^{\varepsilon})_+] > P(B_-) - Q[(B^{\varepsilon})_-]$. Alors

$$(B_S)_+ = B_+ \quad \text{et} \quad P(B_S) - Q[(B_S)^{\varepsilon}] = 2[P(B_S)_+ - Q[(B_S)^{\varepsilon}]_+]$$

$$= 2[P(B_+) - Q[(B_S)^{\varepsilon}]_+] > 2\,[P(B_+) - Q(B^{\varepsilon})_+]$$

d'après l'inclusion signalée plus haut. On en déduit que

$$P(B_S) - Q[(B_S)^{\varepsilon}] > P(B) - Q(B^{\varepsilon})$$

(Si $P(B_S \cup \{0\}) - Q[(B_S \cup \{0\})^{\varepsilon}] > P(B_S) - Q[(B_S)^{\varepsilon}]$

$B_S \cup \{0\}$ est un cas limite d'intervalle symétrique)

52. - LEMME. - Soient P et Q deux lois de probabilité définies par les densités f et g symétriques par rapport à 0 et telles que f et g sont strictement décroissantes sur $\mathbb{R}^+$ avec un seul point d'intersection p. On suppose de plus que sur $[-p, p]$, $f(x) > g(x)$. Alors pour tout borélien B, il existe un intervalle symétrique $[-b, +b]$ tel que

$$\varepsilon[-b, +b] \geq \varepsilon(B)$$

Soit B un borélien de $\mathbb{R}$. D'après le lemme 51, $\exists\, B_S$ borélien symétrique tel que $\varepsilon(B_S) \geq \varepsilon(B)$. Posons $\varepsilon(B_S) = \varepsilon_0$. Alors $[(B_S)^{\varepsilon_0}]_+$ est un ouvert qui peut s'écrire comme union d'intervalles $]a_i, b_i[$ disjoints avec a_i et $b_i \geq 0$ et de longueur $\geq 2\,\varepsilon_0$.

On en déduit que pour tout $\varepsilon < \varepsilon_0$.

$$(B_S)_+ \subset \bigcup_i\,]a_i + \varepsilon, b_i - \varepsilon[\,.$$

$$\text{Soit } b_m = \begin{cases} \text{si } \exists\, i_0 \text{ tel que } p \in\,]a_{i_0} + \varepsilon, b_{i_0} - \varepsilon[\\ \qquad \text{alors } b_m = b_{i_0} - \varepsilon \\ \text{sinon } b_m = \sup\{\, b_i - \varepsilon / b_i - \varepsilon \leq p \,\} \end{cases}$$

Alors on vérifie que

$$P(B_S) - Q(B_S^{\varepsilon}) < P[-b_m, +b_m] - Q[-b_m - \varepsilon, b_m + \varepsilon]$$

Par passage à la limite ($\varepsilon \to \varepsilon_0$) on en déduit le résultat.

53. - LEMME. - En reprenant les hypothèses du lemme 52, soit b_0 le réel > 0 tel que :

$$\varepsilon_1 [0, b_0] = \inf \{\varepsilon / P[0, b] \leq Q[0, b+\varepsilon] + \varepsilon/2\}$$

Alors $\quad \Pi(P, Q) = \varepsilon_1 [0, b_0]$

D'après le lemme 51, on peut se limiter aux intervalles symétriques pour le calcul de Π.

Soit $[-b, b]$ un intervalle symétrique. Alors

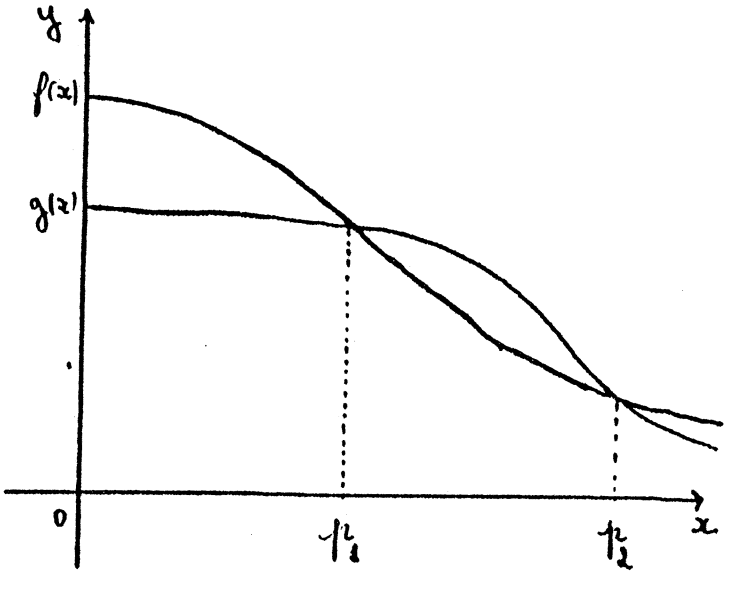
$$P[-b, +b] - Q[-b-\varepsilon, b+\varepsilon] = 2(P[0, b] - Q[0, b+\varepsilon]) \leq \varepsilon_1[0, b_0]$$

54. - LEMME. - Soient P et Q deux probabilités portées par $\mathbb{R}^+$ et définies par les densités f et g répondant au graphe ci-dessous : f et g sont décroissantes, avec deux points d'intersection p_1 et p_2 et sur $[0, p_1[\cup]p_2 + \infty[$ on a : $f(x) > g(x)$. Soit B un borélien de $\mathbb{R}^+$. Alors

$\exists\ b_{m_1} \geq p_1$ et $a_{m_2} \leq p_2$ tel que

$$\varepsilon(B) \leq \varepsilon([0, b_{m_1}] \cup [a_{m_2}, +\infty[)$$

(remarque: $b_{m_1} < a_{m_2}$)



Soit B un borélien de $\mathbb{R}^+$ et $\varepsilon(B) = \varepsilon_1$

Soit alors $\varepsilon < \varepsilon_1$. B^{ε_1} s'écrit comme réunion $]a_i, b_i[$ d'intervalles disjoints de longueur $\geq \varepsilon_1$ et $B \subset \bigcup_i]a_i + \varepsilon, b_i - \varepsilon[$

Posons

$$b_{m_1} = \begin{cases} \text{si } \exists\, i_1 \text{ tel que } p_1 \in]a_{i_1} + \varepsilon, b_{i_1} - \varepsilon[\text{ alors} \\ \qquad b_{m_1} = b_{i_1} - \varepsilon \\ \text{sinon } b_{m_1} = \sup \{ b_i - \varepsilon \,/\, b_i - \varepsilon \leq p_1 \} \end{cases}$$

$$a_{m_2} = \begin{cases} \text{si } \exists\, i_2 \neq i_1 \text{ tel que } p_2 \in]a_{i_2} + \varepsilon, b_{i_2} - \varepsilon[\\ \qquad a_{m_2} = a_{i_2} + \varepsilon \\ \text{sinon } a_{m_2} = \inf \{a_i + \varepsilon \,/\, a_i + \varepsilon \geq p_2\} \end{cases}$$

alors

$$P(B) - Q(B^{\varepsilon_1}) \leq P[0, b_{m_1}] + P[a_{m_2}, +\infty[$$
$$- Q[0, b_{m_1} + \varepsilon[- Q[a_{m_2} - \varepsilon, +\infty[$$

en faisant tendre ε vers ε_1 on en conclut le résultat.

Remarquons enfin que si $\exists\, i_0$ tel que p_1 et $p_2 \in]a_{i_0} + \varepsilon, b_{i_0} - \varepsilon[$

alors $\quad P(B) - Q(B^{\varepsilon_1}) \leq P(\bigcup_i]a_i + \varepsilon, b_i - \varepsilon[) - Q(\bigcup_i]a_i, b_i[)$

$$= P[p_1, p_2] - Q[p_1, p_2] + P[\bigcup_{i \neq i_0}]a_i + \varepsilon, b_i - \varepsilon[]$$

$$- Q(\bigcup_{i \neq i_0}]a_i, b_i[) + P(]a_{i_0} + \varepsilon, p_1[) + P(]p_2, b_{i_0} - \varepsilon[)$$

$$- Q(]a_{i_0}, p_1[) - Q(]p_2, b_{i_0}[)$$

$\leq P[p_1, p_2] - Q[p_1, p_2] + P(D) - Q(D)$

où $D \subset [o, p_1] \cup [p_2, +\infty[$

$\Rightarrow P(B) - Q(B^{\varepsilon_1}) \leq 0$. Ce qui termine la démonstration.

BIBLIOGRAPHIE du paragraphe I. - 1

[1] ABRAMOV V.A. - Estimates for the Levy-Prokhorov distance. Theory of Proba. and its app. (1976), vol. 21, n°2, pp. 396-399.

[2] BLEUEZ - Nouvelle démonstration et extension du théorème de Strassen. Pub. IRMA - Lille, vol. 3, fasc. 6, 1981, pp. 1-21.

[3] BRETAGNOLLE - HUBER C. - Lois empiriques et distance de Prokhorov. Séminaire de Proba. XII. Strasbourg 76/77 (1978) Springer-Verlag, n°649.

[4] CHOQUET G. - Cours d'analyse, tome II, Topologie, Masson (1969).

[5] DUDLEY R.M. - Distances of probability measures and random variables. (1962) A.M.S., vol. 39, n°5, pp. 1563-72.

[6] DUDLEY R.M. - Probabilités and metrics. Lecture notes n°45. Harhus Universitat (1976).

[7] DACUNHA-CASTELLE - Théorie de la robustesse et estimation d'un paramètre. (1977), Astérisque (S.M.F.) n°43-44.

[8] FORD et FULKERSON - (Algorithme de) dans Simonnard . Programmation linéaire. Dunod 1962.

[9] GARCIA-PALOMARES - GINE E. - On the linear programming approach to the optimality property of Prokhorov's distance. (1977) J. of math. analysis and applications, n°60, pp. 596-600.

[10] HAMPEL F.R. - A general qualitative definition of robustness. (1971) A.M.S., vol. 42, n°6, pp. 1887-96.

[11] HUBER P. - Robust statistical procedures (1977) S.I.A.M.

[12] JACOB P. - Convergence faible uniforme de mesures. (1979), C.R.A.S. Série A, T. 273, pp. 1309-11.

[13] PROKHOROV Y.V. - Convergence of random processes and limit theorems in probability theory. (1956) Theory of Proba. and its appl., vol. I, n°2, pp. 157-214.

[14] REVESZ P. - Limit theorems of probability theory. (1975) North. Holland P.C., 420 pages.

[15] REY W. - Robust statistical methods. (1978) Springer-Verlag L.N. in math., vol. 690, 128 pages.

[16] REY W. - Metrics for probability distributions. (1981) Communication personnelle.

[17] SCHAY G. - Nearest random variables with given distributions. (1974) A. of Proba., vol. 2, n°1, pp. 163-6.

[18] STRASSEN V. - The existence of probability measures with given marginals. (1965) A.M.S., vol. 36, pp. 423-39.

[19] SULZGA A. - On minimal metrics in the space of random variables. 13th. E.M.S. Wroclaw 1981.

[20] WELLNER J.A. - A glivenlso Cantelli theorem for empirical measures of independant non identically distributed random variables. (1979). University of Rochester.

[21] YURINSKI V.V. - A smoothing inequality of the Lévy-Prokhorov distance. (1975) Theory of Proba. and its app., vol. XX, n°1, pp. 1-10.

[22] ZOLOTAREV V.M. - On the continuity of stochastic sequences generated by recurrent processes. (1975) Theory of Proba. and its app., vol. XX, n°4, pp. 819-32.

I - 2 Quelques propriétés et calcul de la distance de Geffroy

Introduite par Geffroy et étudiée dans Geffroy-Quidel [1], cette distance s'applique aussi bien aux répartitions ponctuelles qu'aux mesures positives, finies sur tout borélien borné d'un expace vectoriel métrique séparable.

On s'intéresse à diversesmajorations, équivalences et au calcul de cette distance qui se fera, en général, de façon approchée. Divers prolongements sont possibles, en particulier la recherche de la distance sur les variables aléatoires dont la distance de Geffroy est la distance minimale, d'après le concept introduit par Zolotarev.

I - 2-1 Définition - Majoration de la distance de Geffroy.

Soit $(\mathcal{X}, ||.||)$ un espace vectoriel séparable muni de la norme $||.||$ et de sa tribu borélienne $\mathcal{B}$.

1 - Définition : on appelle répartition ponctuelle sur $\mathcal{X}$ toute application $f : \mathcal{X} \to \mathbb{N}$ telle que $S_f = \{x/x \in \mathcal{X}, f(x) > 0\}$ est fini ou dénombrable.

2 - Définition : Soit $A \subset \mathcal{X}$. On appelle effectif de f sur A, l'entier ≥ 0 noté $N(f,A) = \sum_{x \in A} f(x)$. On dira que f est fini si $N(f,\mathcal{X}) < +\infty$.

De plus si $A = \emptyset$ on pose $N(f,A) = 0$. On définit également pour $A \subset \mathcal{X}$ et $\delta > 0$:
$A^{\delta} = \{x \in \mathcal{X} / \exists\, a \in A, ||x-a|| < \delta\}$ et $A^{\delta]} = \{x \in \mathcal{X} / \exists\, a \in A, ||x-a|| \leq \delta\}$.
On pose $\emptyset^{\delta} = \emptyset$.

On notera $b(0,r)$ la boule ouverte de centre 0 et de rayon r et $\hat{b}(0,r)$ la boule fermée. On appelle $\mathcal{F}$ l'ensemble des répartitions pontuelles sur $\mathcal{X}$ dont les effectifs sont finis sur tout borélien borné.

3 - <u>Définition</u> : soient f et g deux éléments quelconques de $\mathcal{F}$ et $r \in \mathbb{R}^{+*}$. on définit

(4) $$\sigma_r(f,g) = \inf \{\delta > 0 \ / \ \forall A \in \mathcal{B} \text{ tel que}$$

$$A^\delta \subset b(0,r) : N(f,A) \leq N(g,A^\delta) + \delta\}$$

puis

(5) $$\rho_r(f,g) = \max(\sigma_r(f,g) \ ; \ \sigma_r(g,f))$$

la fonction ρ_r ainsi définie est ≥ 0 et vérifie l'inégalité triangulaire.

6 - <u>Remarque</u> : on ne change pas la valeur de $\rho_r(f,g)$ en remplaçant $b(0,r)$ par $\hat{b}(0,r)$. En effet on a $A^\delta \subset b(0,r) \Longleftrightarrow A^\delta \subset \hat{b}(0,r)$

7 - <u>Lemme</u> : <u>On ne change pas la valeur de</u> $\sigma_r(f,g)$ <u>en remplaçant</u> A^δ <u>par</u> $A^{\delta]}$ <u>dans la formule (4)</u>.

Preuve : notons $\sigma_r^{]}(f,g) = \inf \{ \delta > 0 \ / \ \forall A \in \mathcal{B}$ tel que $A^{\delta]} \subset \hat{b}(0,r) : N(f,A) \leq N(g,A^{\delta]}) + \delta\}$.

D'après la remarque 6 on sait que $\sigma_r(f,g) = \inf\{\delta > 0 \ / \ \forall A \in \mathcal{B}$ et $A^\delta \subset \hat{b}(0,r) : N(f,A) \leq N(g,A^\delta) + \delta\}$. Alors la relation : $\left(A^\delta \subset b(0,r)\right) \Leftrightarrow \left(A^{\delta]} \subset \hat{b}(0,r)\right)$ et l'inégalité $N(g,A^\delta) \leq N(g,A^{\delta]})$ entrainent : $\sigma_r^{]} \leq \sigma_r$.

Réciproquement soit A et δ_0 tels que : $A^{\delta_0]} \subset \hat{b}(0,r)$ avec $N(f,A) > N(g,A^{\delta_0})+\delta_0$ et $N(f,A) \leq N(g, A^{\delta_0]})+ \delta_0$. On voit que :

$$\inf\{\delta>0 \ / \ N(f,A) \leq N(g,A^{\delta]})+\delta\} = \inf\{\delta>0 \ / \ N(f,A) \leq N(g,A^\delta)+\delta\} = \delta_0$$

On en déduit que $\sigma_r^{]} \geq \sigma_r$.

8 - <u>Remarque</u> : si $\sigma_r(f,g) = \varepsilon_0$ alors $\forall \varepsilon < \varepsilon_0$ et $\forall A$ tel que $A^\varepsilon \subset b(0,r)$ on a : $N(f,A) < N(g,A^\varepsilon) + \varepsilon$.

Ceci résulte de la croissance stricte de l'application $\varepsilon \longrightarrow N(g,A^\varepsilon) +\varepsilon$ et de la définition de σ_r. L'inégalité même au sens large est en général fausse pour $\varepsilon = \varepsilon_0$ comme on peut le voir sur l'exemple suivant.

9 - Exemple : $\mathcal{X} = \mathbb{R}$, $S_f \cap [0,2] = \{3/4\}$

$S_g \cap [0,2] = \{1\}$.

$f(3/4) = 3$ $\quad g(1) = 2$ et prenons $A = \{3/4\}$.

Alors $\sigma_1(f,g) = 1/4$ et $N(f,A) > N(g, A^{1/4}) + 1/4$.

10 - Lemme : Soit A un borélien de $\mathcal{X}$, $A \subset B(0,r)$.

Soit $\Delta = \{\delta > 0 \ / \ A^{\delta} \subset b(0,r)\} \neq \emptyset$. Supposons de plus que $\forall\ \delta \in \Delta$, $N(f,A) > N(g, A^{\delta}) + \delta$. Alors $\sigma_r(f,g) \geq \sup\Delta$.

ceci résulte immédiatement de la définition 4.

11 - Définition : Soient f et g deux éléments de $\mathcal{F}$. On appelle distance de Geffroy l'application $\rho : \mathcal{F} \times \mathcal{F} \longrightarrow \mathbb{R}^+$ définie par :

$$(f,g) \longrightarrow \rho(f,g) = \sum_{m=1}^{+\infty} 2^{-m} \frac{\rho_m(f,g)}{1+\rho_m(f,g)}$$

-la distance de Geffroy est comprise entre 0 et 1.

de $\quad 0 \leq \dfrac{\rho_m(f,g)}{1+\rho_m(f,g)} \leq 1 \quad$ on déduit

$$0 \leq \sum_{m=1}^{+\infty} 2^{-m} \frac{\rho_m}{1+\rho_m} \leq \sum_{m=1}^{+\infty} 2^{-m} = \frac{1}{2}\left(\sum_{0}^{+\infty} \left(\frac{1}{2}\right)^m\right) = \frac{1}{2}\left(\frac{1}{1-1/2}\right) = 1.$$

12 - Définition : nous dirons que l'espace vectoriel normé $(\mathcal{X}, ||.||)$ vérifie la propriété (p_1) si et seulement si :

$$(p_1) \quad \forall \quad 0 \leq r \leq r' \ : b(0,r) \not\subset b(0,r').$$

On en déduit facilement que $\forall\ x \in \mathcal{X}$ $b(x,r) \not\subset b(x,r')$ et que $\forall\ A \subset \mathcal{X}$, $A^r \not\subset A^{r'}$

13 - Lemme : Soit $(\mathcal{X}, ||.||)$ un espace vectoriel que vérifie la propriété (p_1) ; f et g étant deux élements de $\mathcal{F}$. Alors $\sigma_r(f,g) \leq r$; cette valeur pouvant être atteinte.

Preuve : Soit $r' > r$. D'après la propriété (p_1) $\{A \in \mathcal{B} / A^{r'} \subset b(0,r)\} = \emptyset$ et

l'inégalité de la définition 4 se réduit à $0 \leq 0+r'$. Donc $\sigma_r \leq \inf\{r'/r'>r\} = r$.
D'autre part compte tenu du lemme 10 on voit que la minoration la plus grande de σ_r est obtenue pour $A = \{0\}$, $\Delta = [0,r]$ et $N(f,A) > N(g,A^\delta) + \delta \ \forall \ \delta \in \Delta$.
Ce qui conduit à $\sigma_r(f,g) \geq r$; soit $\sigma_r(f,g) = r$ dans ce cas.

14 - <u>Lemme</u> : <u>Soit</u> $(\mathcal{X}, ||.||)$ <u>un espace vectoriel séparable qui vérifie la propriété</u> (p_1). <u>Alors la distance de Geffroy prend ses valeurs dans</u> $[0, 2(1-\text{Log } 2)]$.

preuve : d'après le lemme 13 $\rho(f,g) = \sum_1^{+\infty} 2^{-m} \frac{\rho_m}{1+\rho_m} \leq \sum_1^{+\infty} 2^{-m} \frac{m}{1+m} =$

$$\sum_1^{+\infty} 2^{-m} - \sum_1^{+\infty} \frac{2^{-m}}{m+1} = 1 - 2 \sum_1^{+\infty} \frac{2^{-m+1}}{m+1}$$

Si $|x| < 1$ on a $\frac{1}{1-x} = 1 + x + \ldots + x^m + \ldots$. Par intégration on obtient

$$\int_0^x \frac{du}{1-u} = \int_0^x (1 + u + \ldots + u^m + \ldots)\, du = \sum_0^{+\infty} \int_0^x u^m\, du = \sum_0^{+\infty} \frac{x^{-m+1}}{m+1} = \text{Log} \frac{1}{|1-x|}.$$

En posant $x = 1/2$ on obtient

$$\text{Log } 2 - \frac{1}{2} = \sum_1^{+\infty} \frac{2^{-m+1}}{m+1} \quad \text{d'où l'on déduit} \quad \rho \leq 1-2(\text{Log } 2 - 1/2)$$

On suppose par la suite que l'on travaille sur un espace vectoriel séparable et normé $(\mathcal{X}, ||.||)$ qui vérifie la propriété (p_1) : c'est le cas des espaces $\mathbb{R}$, $\mathbb{R}^p$.

Soit $0 \leq \alpha \leq r$ fixé. Posons :

(15) $$\beta_r(\alpha,f,g) = \inf \{\beta > 0 \ / \ \forall \ A \in \mathcal{B} \text{ tel que } A^{\alpha]} \subset \overset{\frown}{b}(0,r) :$$

$$N(f,A) \leq N(g,A^{\alpha]}) + \beta\}.$$

On voit que $A^{\alpha]} \subset \overset{\frown}{b}(0,r) \Longleftrightarrow A \subset \overset{\frown}{b}(0,r-\alpha)$.

16 - <u>Lemme</u> : <u>l'application</u> $\alpha \to \beta_r(\alpha,f,g)$ <u>est une fonction non croissante de</u> α <u>de</u> $[0,r]$ <u>dans</u> $[0,r]$.

Preuve : si $\alpha_1 < \alpha_2$ dans $[0,r]$, $A^{\alpha_1} \not\subset A^{\alpha_2} \Rightarrow \{A \in \mathcal{B} / A^{\alpha_2]} \subset b(0,r)\} \not\subset \{A \in \mathcal{B} / A^{\alpha_1]} \subset \hat{b}(0,r)\}$.

D'autre part soit $A^{\alpha_1]} \subset \hat{b}(0,r)$ tel que $N(f,A) \leq N(g,A^{\alpha_1]}) + \alpha_1$ et tel que $A^{\alpha_2]} \subset \hat{b}(0,r)$. Alors $N(g,A^{\alpha_1]}) \leq N(g,A^{\alpha_2]}) \Rightarrow N(f,A) < N(g,A^{\alpha_2]}) + \alpha_2$.

17 - <u>Lemme</u> : $\sigma_r(f,g) = \inf \{\alpha / \alpha \in [0,r] , \beta_r(\alpha,f,g) \leq \alpha\}$

Preuve : l'inégalité $\beta_r(\alpha,f,g) \leq \alpha$ est moins vérifiée pour $\alpha \leq r$. D'autre part si $\beta_r(\alpha,f,g) \leq \alpha$ on voit que $\sigma_r(f,g) \leq \alpha$ d'après la non croissance de la fonction $\alpha \longrightarrow \beta_r(\alpha,.,.)$. Enfin supposons $\sigma_r < \inf\{\alpha / \beta_r(\alpha,f,g) \leq \alpha\}$.
Soit $\sigma_r < \alpha_0 < \inf \{\alpha / \beta_r(\alpha,f,g) \leq \alpha\}$. Alors $\forall$ A tel que $A^{\alpha_0]} \subset \hat{b}(0,r)$ on a $N(f,A) \leq N(g,A^{\alpha_0]}) + \sigma_r$ donc $\alpha_0 \geq \inf\{\alpha / \beta_r(\alpha,f,g) \leq \alpha\}$. D'où la contradiction.

I - 2 - 2 <u>Calcul de la distance de Geffroy</u>

La distance de Geffroy est obtenue en faisant la somme d'une série. On va d'abord faire une majoration de l'erreur de troncature pour savoir combien de termes il faudra calculer puis additionner.

Soit $0 < \varepsilon < 1$ donné.

$$\rho(f,g) - \sum_{1}^{m_0} 2^{-m} \frac{\rho_m(f,g)}{1+\rho_m(f,g)} = \sum_{m_0+1}^{+\infty} 2^{-m} \frac{\rho_m}{1+\rho_m} < \sum_{m_0+1}^{+\infty} 2^{-m} = 2^{-m_0}.$$

$2^{-m_0} < \varepsilon \Leftrightarrow 2^{m_0} > \frac{1}{\varepsilon}$ soit $m_0 > - \frac{\text{Log } \varepsilon}{\text{Log } 2}$ par exemple $m_0 = [- \frac{\text{Log } \varepsilon}{\text{Log } 2}] + 1$.

Compte tenu du lemme 14, on pourra si nécessaire obtenir une majoration plus fine de l'erreur de troncature qui permettra une économie de temps de calcul.

Pour le calcul de $\sum_{1}^{m_0} 2^{-m} \frac{\rho_m}{1+\rho_m}$ on est ramené au calcul de $\sigma_m(f,g)$ avec m entier $\in \{1,\ldots,m_0\}$.

Soit $\alpha \in [0,m]$ fixé. Posons

$S_f \cap \hat{b}(0,m-\alpha) = \{x_1,\dots,x_l\} = S_{f,m-\alpha}$

$S_g \cap \hat{b}(0,m) = y_1,\dots,y_n = S_{g,m}$

puis $f(x_i) = p_i$ pour $i = 1,\dots,l$

$g(y_j) = q_j$ pour $j = 1,\dots,n$.

Soit (u,v) un élément de $\mathbb{R}^{l+n}$ de coordonnées $(u_i,\ i = 1,\dots,l\ ;\ v_j,\ j = 1,\dots,n)$.

Posons $$d_{ij} = \begin{cases} 1 & \text{si } ||x_i - y_j|| \leq \alpha \\ 0 & \text{si } ||x_i - y_j|| > \alpha \end{cases}$$

Considérons la forme liléaire sur $\mathbb{R}^{l+n}$

$$W(u,v) = \sum_{i=1}^{l} p_i u_i + \sum_{j=1}^{n} q_j v_j .$$

18 - <u>Lemme</u> : <u>l'égalité suivante est vérifiée</u> :

$$\beta_m(\alpha,f,g) = N(f,\hat{b}(0,m-\alpha)) - \min_{(u,v)} W(u,v)$$

<u>où le minimum de la forme linéaire W est obtenue sous les contraintes</u> :

$$(19) \quad \left.\begin{array}{l} \forall\ i = 1,\dots,l \\ \forall\ j = 1,\dots,n \end{array}\right\} \quad u_i \geq 0 \ ; \ v_j \geq 0 \qquad u_i + v_j \geq d_{ij}$$

Preuve : on sait que le minimum de la forme linéaire W sous les contraintes (19) est obtenue sur l'un des sommets du polyèdre convexe engendré par les contraintes et donc ici l'un des sommets du cube unité de $\mathbb{R}^{l+n}$. Un tel sommet est caractérisé par un couple (u,v) dont les coordonnées valent 0 ou 1. A toute suite (u_i) $i = 1,\dots,l$ de 0 ou de 1 est associée de façon unique un sous ensemble B de $S_{f,m-\alpha}$ tel que $u_i = 0 \iff x_i \in B$

$u_i = 1 \iff x_i \in S_{f,m-\alpha} - B$.

Dans ce cas on a :

$d_{ij} - u_i = 1$ si $||y_j - x_i|| \leq \alpha$ et $u_i = 0$

$$d_{ij} - u_i = 0 \quad \text{si} \quad \begin{cases} ||y_j - x_i|| \leq \alpha \quad \text{et } u_i = 1 \\ \text{ou} \\ ||y_j - x_i|| > \alpha \quad \text{et } u_i = 0 \end{cases}$$

$$d_{ij} - u_i = -1 \text{ si} \quad ||y_j - x_i|| > \alpha \quad \text{et } u_i = 1$$

Les valeurs v_j qui respectent les contraintes (19) sont données par :

$$v_j \geq \begin{cases} 1 & \text{si} \quad y_j \in B^{\alpha]} \cap S_g \\ 0 & \text{si} \quad y_j \in S_g \cap (\hat{b}(0,m) - B^{\alpha]}) \end{cases}$$

Donc la suite (u,v) définie par

$$u_i = 0 \quad \text{si} \quad x_i \in B \quad ; \quad u_i = 1 \quad \text{si} \quad x_i \in S_{f,m-\alpha} - B$$

$$v_j = 0 \quad \text{si} \quad y_j \in S_g \cap (\hat{b}(0,m) - B^{\alpha]}) \quad ; \quad v_j = 1 \quad \text{si} \quad y_j \in B^{\alpha]} \cap S_g$$

caractérise un sommet du cube unité de $\mathbb{R}^{1+n}$ situé dans la zone des contraintes. Pour ce sommet on a :

$$\sum_{i=1}^{1} p_i u_i + \sum_{j=1}^{n} q_j v_j = N(f, \hat{b}(0,m-\alpha) - B) + N(g, B^{\alpha]})$$

D'autre part si x est un autre sommet du cube unité vérifiant les contraintes (19) tel que $u_i = 1$ si $x_i \in \hat{b}(0,m-\alpha) - B$ on a $W(x_B) \leq W(x)$ car on a choisi pour la suite v_j les plus petites valeurs vérifiant les contraintes. Donc :

$$\min W(u,v) = \min_{B \subset S_{f,m-\alpha}} \left[N(f, S_{f,m-\alpha} - B) + N(g, S_g \cap B^{\alpha]})\right]$$

$$= N(f, S_{f,m-\alpha}) - \max_{B \subset S_{f,m-\alpha}} \left[N(f,B) - N(g, S_g \cap B^{\alpha]})\right]$$

$$= \inf \{ \beta / \forall B \subset S_{f,m-\alpha}, N(f,B) \leq N(g, B^{\alpha]}) + \beta\}$$

$$= \inf \{ \beta / \forall A \text{ tel que } A^{\alpha]} \subset \hat{b}(0,m) \text{ on a : } N(f,A) \leq N(g, A^{\alpha]}) + \beta\}$$

en remarquant que si $A^{\alpha]} \subset \widehat{b}(0,m)$

$N(f,A) = N(f, A \cap S_{f,m-\alpha})$ et

$N(g,A^{\alpha]}) = N(g, S_g \cap [A \cap S_{f,m-\alpha}]^{\alpha]})$.

Soit $B \subset S_{f,m}$. $B = \{x_{i_1},\dots, x_{i_p}\}$ et tel que $B^{\alpha]} \subset b(0,m)$ pour $\alpha \in [\alpha_0, \alpha_1]$

L'application $\alpha \to N(g, B^{\alpha]})$ a des sauts pour des $\alpha \in \{||y_j-x_{i_k}|| \ / \ k=1,\dots,p$ $j=1,\dots n\}$.

On va considérer $E_m = \{ ||x_i-y_j|| / \ x_i \in S_{f,m} , y_j \in S_{g,m} , ||x_i-y_j|| \le m\}$

Les valeurs des α testés appartiendront à cet ensemble. Ici pourra être mis au point un algorithme de solution rapide : l'ordinateur fera le calcul des $\beta_m(\alpha,f,g)$ au moyen de l'algorithme de Ford et Fulkerson. On veut limiter au maximum le nombre de valeurs $\alpha \in [0,m]$ pour lesquelles on fera ce calcul et limiter le temps calcul de nouvelles valeurs de β_m à partir de valeurs déjà connues.

Par exemple, si l'on choisi de tester $\alpha_0 \in [0,m]$ on énumère $S_f \cap \widehat{b}(0, m-\alpha_0)$ et si

$\alpha_0 \notin \{ ||y_j-x_i|| \ / \ x_i \in S_{f,m-\alpha}, y_j \in S_{g,m} \}$,

on passe à la plus grande valeur α_1 inférieure à α_0 qui appartient à cet ensemble et $\beta_m(\alpha_0,f,g) = \beta_m(\alpha_1,f,g)$.

D'autre part, si l'on a testé α_0 et que l'on teste $\alpha_1 < \alpha_0$ et que $S_{f,m-\alpha_1} = S_{f,m-\alpha_0}$ on a $\beta_m(\alpha_0,f,g) = \beta_m(\alpha_1,f,g)$.

A l'aide de l'ordinateur on va trouver

$\alpha_{min} = \min\{ \alpha / \alpha \in E_m$ tel que $\beta_m(\alpha,f,g) \le \alpha\}$

$\alpha_{max} = \max\{ \alpha / \alpha \in E_m$ tel que $\beta_m(\alpha,f,g) > \alpha\}$

(19) Alors : $\sigma_m(f,g) = \min(\alpha_{min} ; \beta_m(\alpha_{max},f,g))$.

Des résultats complémentaires sur σ_m peuvent être obtenus lorsque $S_{f,m}$ et $S_{g,m}$ ont une intersection non vide ou lorsque $\mathcal{X} = \mathbb{R}$.

On passe alors au calcul de $\sigma_m(g,f)$ pour en conclure $\rho_m(f,g)$ puis

$$\sum_{m=1}^{m_0} 2^{-m} \frac{\rho_m(f,g)}{1+\rho_m(f,g)}$$

BIBLIOGRAPHIE du paragraphe I - 2

J. Geffroy - P. Quidel : convergences stochastiques des répartitions ponctuelles aléatoires.
Révista da fac. de Ciên. da Univ. de Coimbra
vol XLIX, 1974 p. 1-54

I-3 LES DISTANCES DE FORTET-MOURIER ET DE WASSERSTEIN

I-3.1 Introduction.

On s'intéresse maintenant à diverses distances entre mesures ou probabilités qui ont été introduites durant les trente dernières années. Il s'agit tout d'abord de la distance de Fortet-Mourier qui a été proposée en 1953 et d'une variante qui a surtout été étudiée par R. Dudley. Ces deux distances ont une importance croissante. Dudley a en effet montré que la variante est l'une des distances associées à la topologie de la convergence faible et qu'elle est équivalente à la distance de Prokhorov. On verra sur des exemples que les inégalités obtenues sont optimales. Signalons qu'une seconde variante, également équivalente à la distance de Prokhorov a été récemment introduite dans son livre par Huber.

Il se trouve, d'autre part, que la distance de Fortet-Mourier est égale à une autre distance définie à l'aide de probabilités sur l'espace produit dont les marginales sont données, appelée distance de Wasserstein. C'est l'objet d'un théorème qui a été récemment démontré de façon directe (1982) et qui porte le nom de théorème de Kantorovitch et Rubinstein. Ces résultats illustrent la dualité entre l'espace des fonctions lipschitziennes de rapport 1 et l'espace des mesures signées finies par rapport auxquelles une condition d'intégrabilité est vérifiée.

Après quelques rappels sur l'algèbre des fonctions lipschitziennes, quelques définitions et quelques exemples, on s'intéresse au calcul de la distance de Fortet-Mourier. On se ramène par discrétisation à un problème de transport connu sous le nom de problème de Hitchcock.

Puis on s'intéresse à un aspect du théorème de Kantorovitch et Rubinstein: au lieu de raisonner uniquement sur les lois de probabilités P et Q on s'intéresse également aux variables aléatoires X et Y

qui ont pour lois respectives P et Q .

On démontre en particulier que si P et Q sont à support fini, alors pour toute variable aléatoire X de loi P , il existe une variable aléatoire Y de loi Q telle que la distance de Fortet-Mourier entre P et Q est égale à l'espérance de $d(X,Y)$. Ce qui veut dire que la distance de Wasserstein est alors un minimum atteint pour X et Y .

I.3.2. <u>Généralités</u>.

Soit (S,d) un espace polonais muni de sa tribu des boréliens $\mathcal{B}$. On appellera $\mathcal{P}(S)$ l'ensemble des mesures de probabilités de Borel sur $(S,\mathcal{B})$. On notera :

$$\mathcal{M}(S) = \{\mu / \text{mesure signée finie sur } S\}$$

$$\mathcal{M}^+(S) = \{\mu \in \mathcal{M}(S) \ / \ \mu \text{ non négative}\}$$

$$\mathcal{M}_1^+(S) = \{\mu \in \mathcal{M}(S) / \ \exists x : \int d(x,y)\, d\mu(y) < +\infty\}$$

$$\mathcal{M}_1(S) = \{\mu \in \mathcal{M}(S) / \ \mu \in \mathcal{M}_1^+(S)\}$$

$$\mathcal{P}_1(S) = \{P \in \mathcal{P}(S) / \ \exists x \in S : \int d(x,y)\, dP(y) < +\infty\}$$. On a donc :

$$\mathcal{P}_1(S) = \mathcal{P}(S) \cap \mathcal{M}_1^+(S) .$$

En utilisant l'inégalité triangulaire on voit facilement que si $\exists x \in S$ tel que $\int_S d(x,y)\, dP(y) < +\infty$, cette propriété est vérifiée pour tout $x \in S$.
D'autre part si S est borné $\mathcal{P}_1(S) = \mathcal{P}(S)$.

Soit une fonction $f : S \longrightarrow \mathbb{R}$. On définit

$$(1) \qquad \|f\|_L = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)} .$$

Si $f = cte$ $\|f\|_L = 0$ et donc $\|f\|_L$ n'est qu'une semi-norme. On pose

(2) $$\|f\|_\infty = \sup_{x \in S} |f(x)|$$

puis

(3) $$\|f\|_{BL} = \|f\|_L + \|f\|_\infty .$$

Comme $\|f\|_\infty$ est une norme et $\|f\|_L$ une semi-norme $\|f\|_{BL}$ est une norme. On appellera

(4) $$BL(S,d) = \{ f : S \longrightarrow \mathbb{R} \,/\, \|f\|_{BL} < +\infty \} .$$

Alors $(BL(S,d) , \|\cdot\|_{BL})$ est une algèbre de Banach commutative et unitaire et l'on a :

$$\|f.g\|_{BL} \leq \|f\|_{BL} \cdot \|g\|_{BL} .$$

D'autre part si l'on note $\max(f,g) = f \vee g$ et $\min(f,g) = f \wedge g$ on a :

$$\|f_1 \vee f_2 \vee \ldots \vee f_n\|_L \leq \max_{i=1,\ldots,n} \|f_i\|_L$$

$$\|f_1 \vee f_2 \vee \ldots \vee f_n\|_{BL} \leq 2 \max_{i=1,\ldots,n} \|f_i\|_{BL}$$

de même $\|f_1 \wedge f_2 \wedge \ldots \wedge f_n\|_L \leq \max_{i=1\ldots n} \|f_i\|_L$ et

$\|f_1 \wedge f_2 \wedge \ldots \wedge f_n\|_{BL} \leq 2 \max_{i=1\ldots n} \|f_i\|_{BL}$.

Si P et Q sont deux éléments de $\mathcal{P}(S)$ (et plus généralement de $\mathcal{M}(S)$) on peut définir

(5) $$d_{BL}(P,Q) = \sup \{ \left|\int f\, d(P-Q)\right| / \|f\|_{BL} \leq 1 \} .$$

Si P et $Q \in \mathcal{P}_1(S)$ on peut définir

(6) $$\|P-Q\|_L^* = \sup \{ \left|\int f\, d(P-Q)\right| / f : \|f\|_L \leq 1 \} .$$

Plus généralement si $\mu \in \mathcal{M}_1(S)$ on peut définir

$$\|\mu\|_L^* = \sup \{ |\int f d\mu / \|f\|_L \leq 1 \} .$$

R. Fortet et E. Mourier ont introduit en 1953 la norme (6). La distance (5) a surtout été étudiée par R. Dudley.

7. <u>Lemme</u> : d_{BL} <u>est une distance sur</u> $\mathcal{P}(S)$.

<u>Preuve</u> :

i) si $P = Q$ il est clair que $d_{BL}(P,Q) = 0$.

Inversement supposons que $d_{BL}(P,Q) = 0$. Soit F un fermé de $\mathcal{B}$. Considérons les fonctions f_n définies par :

$$f_n(x) = \begin{cases} 1 \wedge n\, d(x,F) & \text{si } x \in F^{1/n} \\ 0 & \text{si } x \in S - F^{1/n} \end{cases}$$

où $F^{1/n}$ est le dilaté d'ordre $1/n$ de F .

Alors il est clair que $\|f_n\|_\infty = 1$.

D'autre part si $x \in F$ et $y \in F^{1/n} - F$ on a :

$$|f(x) - f(y)| = n\ d(y,F) \leq n\, d(x,y)$$

si $x \in F^{1/n} - F$ et $y \in F^{1/n} - F$:

$$|f(x) - f(y)| = n\, |d(x,F) - d(y,F)| \leq n\, d(x,y)$$

enfin si $x \in F^{1/n}$ et $y \in \complement F^{1/n}$ on a :

$$|f(x) - f(y)| = 1 - n\ d(x,F)$$

supposons que $d(x,F) = 1/\alpha n \quad \alpha > 1$

alors $d(x,y) > \dfrac{1}{n} - \dfrac{1}{\alpha n} = \dfrac{\alpha - 1}{\alpha n}$

$$|f(x) - f(y)| / d(x,y) \leq \frac{\alpha - 1}{\alpha} \quad \frac{\alpha n}{\alpha - 1} = n .$$

On en déduit que $\|f_n\|_L \leq n$ et $\|f_n\|_{BL} \leq 1 + n$. D'autre part la suite (f_n) décroît vers $1/F$.

Alors on a : $0 \geq \left| \int \frac{1}{n+1} f_n \, d(P,Q) \right| \geq 0 \Longrightarrow \int f_n \, dP = \int f_n \, dQ$.

Par passage à la limite on a : $P(F) = Q(F)$ pour tout F fermé de $\mathcal{B}$. On en déduit que $P = Q$.

ii) il est clair que $\forall f \in BL$

$$\left| \int f \, d(P-Q) \right| = \left| \int f \, d(Q-P) \right| \quad \text{donc}$$

$$d_{BL}(P,Q) = d_{BL}(Q,P) .$$

iii) soient $P, Q, R \in \mathcal{P}(S)$ et $f \in BL(S,d)$

$$\left| \int f \, d(P-Q) \right| \leq \left| \int f \, d(P-R) + \int f \, d(R-Q) \right|$$

$$\leq \left| \int f \, d(P-R) \right| + \left| \int f \, d(R-Q) \right| \quad \text{d'où}$$

$$\sup_{\|f\|_{BL} \leq 1} \left| \int f \, d(P-Q) \right| \leq \sup_{\|f\|_{BL} \leq 1} \left| \int f \, d(P-R) \right| + \sup_{\|f\|_{BL} \leq 1} \left| \int f \, d(R-Q) \right|$$

c.q.f.d.

on a alors le résultat suivant (cf. [7])

8. <u>Théorème</u> : <u>Soit</u> (P_n) <u>une suite de probabilités dans</u> $\mathcal{P}(S)$. <u>Alors</u>

$$P_n \longrightarrow P \text{ faiblement} \Longleftrightarrow d_{BL}(P_n, P) \underset{n \to +\infty}{\longrightarrow} 0 .$$

On sait que la notion de convergence étroite (ou faible) des probabilités est définie à partir de l'ensemble des fonctions continues bornées : $C_b(S,d)$. Le théorème 8 semble indiquer que l'on peut se limiter aux fonctions lipschitziennes bornées. On est alors en droit de se demander s'il n'existe pas un résultat de densité de $BL(S,d)$ dans $C_b(S,d)$; et en effet :

si (S,d) est un espace compact $BL(S,d)$ est dense dans $C_b(S,d)$ pour $\|\cdot\|_\infty$. Si (S,d) est seulement un espace métrique complet $BL(S,d)$ est dense dans l'ensemble des fonctions uniformément continues, bornées (cf. [5]), pour la même norme.

On sait (cf. [7]) qu'il existe une autre distance qui métrise la topologie de la convergence faible dans $\mathcal{P}(S)$: c'est la distance de Prokhorov. On rappelle que π est définie par
$\pi(P,Q) = \inf \{ \varepsilon \,/\, \forall F$ fermé de $\mathcal{B}$: $P(F) \leq Q(F^\varepsilon) + \varepsilon \}$ et $\pi(P,Q) \leq 1$.

On peut alors se demander si ces deux distances sont équivalentes : ce qui n'a rien d'évident car deux distances peuvent être associées à la même topologie et ne pas être équivalentes : ex : sur $\mathbb{R}^+$
$\left|\frac{1}{x} - \frac{1}{y}\right|$ et $|x-y|$ définissent la topologie habituelle et l'on n'a pas
$\left|\frac{1}{x} - \frac{1}{y}\right| \longrightarrow 0$ quand $|x-y| \longrightarrow 0$. Ici cependant :

9. <u>Proposition</u> : <u>On a les inégalités suivantes</u> :

$$\frac{2}{3}\ \pi^2(P,Q) \leq \|P-Q\|^*_{BL} \leq 2\pi(P,Q) .$$

a) supposons qu' $\exists$ α et β tel que

$$P(A) \leq Q(A^{\alpha]}) + \beta \quad \text{pour tout } A \text{ fermé de } \mathcal{B} .$$

Alors d'après un résultat de Dudley [6] $\forall \varepsilon > 0$ $\exists$ une loi $\mu \in \mathcal{P}(S \times S)$ telle que

$$\mu \circ \pi_1^{-1} = P \quad \text{et} \quad \mu \circ \pi_2^{-1} = Q \quad \text{et} \quad \mu(d(x,y) > \alpha+\varepsilon) \leq \beta + \varepsilon .$$

On peut donc trouver des variables aléatoires à valeurs dans (S,d) : X et Y de lois respectives P et Q telles que $P(d(x,y) > \alpha + \epsilon) \leq \beta + \epsilon$.

Soit $f \in BL(S,d)$

$$\left| \int f \, d(P-Q) \right| = \left| E \left[f(x) - f(y) \right] \right|$$

$$= \left| E \left[\{f(x) - f(y)\} \{ \mathbb{1}_{\{d(x,y) \leq \alpha + \epsilon\}} \} + \{f(x) - f(y)\} \{ \mathbb{1}_{\{d(x,y) > \alpha + \epsilon\}} \} \right] \right|$$

$$\leq \left| E \left[\frac{|f(x) - f(y)|}{d(x,y)} (\alpha + \epsilon) + 2 \sup_{x \in S} |f(x)| \, \mathbb{1}_{\{d(x,y) > \alpha + \epsilon\}} \right] \right|$$

$$\leq \|f\|_L (\alpha + \epsilon) + 2 \|f\|_\alpha (\beta + \epsilon)$$

$$\leq 2 \max(\alpha + \epsilon, \beta + \epsilon) (\|f\|_L + \|f\|_\infty)$$

en faisant tendre ϵ vers 0 , sachant que d'après l'hypothèse de départ $\pi(P,Q) \leq \max(\alpha, \beta)$, on conclut que

$$d_{BL}(P,Q) \leq 2\pi(P,Q) .$$

b) Soit A un fermé de S . On définit la fonction qui vaut :

$$f(x) = \begin{cases} 1 & \text{si} \quad x \in A \\ -1 & \text{si} \quad x \in \complement A^\beta \end{cases}$$

avec $\|f\|_\infty = 1$ et $\|f\|_{BL} \leq 1 + 2/\beta$

alors $\|P-Q\|^*_{BL} \geq \dfrac{1}{1+2/\beta} \int f \, d(P-Q)$

$$= \frac{1}{1+2/\beta} \int (f+1) \, d(P-Q) \geq \frac{2}{1+2/\beta} \left[P(A) - Q(A^\beta) \right] \geq 2\alpha / 1+(2/\beta).$$

On en conclut que

$$\| P-Q \|_L^* \geq \frac{2\alpha\beta}{2+\beta} \quad \text{et si} \quad \alpha = \beta = \pi(P,Q)$$

$$\| P-Q \|_L^* \geq \frac{2\pi^2(P,Q)}{2+\pi(P,Q)} \geq \frac{2\pi^2(P,Q)}{3} \qquad \text{c.q.f.d.}$$

<u>Exemples de calcul</u> :

<u>exemple 1</u>. On va supposer que P et Q sont deux masses de Dirac en x_o et y_o respectivement. Alors

$$\| \delta_{x_o} - \delta_{y_o} \|_L^* = \sup \{ \, | f(x_o) - f(y_o) | \; / \; \| f \|_L \leq 1 \, \}$$

donc $| f(x_o) - f(y_o) | \leq d(x_o, y_o)$.

D'autre part, soit $f(x) = d(x, x_o)$; on a : $\| f \|_L \leq 1$ donc

$$\| \delta_{x_o} - \delta_{y_o} \|_L^* \geq | d(x_o, y_o) - d(x_o, x_o) | = d(x_o, y_o) \; .$$

On en conclut que

$$\| \delta_{x_o} - \delta_{y_o} \|_L^* = d(x_o, y_o) \; .$$

<u>exemple 2</u>. Nous allons supposer que $S = \mathbb{R}$ et poser $y_o = x_o + \frac{1}{n}$. Définissons f par

$$f(x) = \begin{cases} 1 & \text{si } x \geq y_o \\ 2n(x-x_o) - 1 & \text{si } x_o < x < y_o \\ -1 & \text{si } x \leq x_o \end{cases}$$

alors $\|f\|_\infty = 1$ et $\|f\|_L = 2n$ donc

$$d_{BL}(\delta_{x_o}, \delta_{y_o}) \geq \frac{1}{1+2n} \left| \int f \, d(\delta_{x_o}, \delta_{y_o}) \right| = \frac{2}{1+2n} .$$

Nous allons alors utiliser le résultat suivant :

10. <u>Lemme</u> : Soient δ_{x_o} et δ_{y_o} deux masses de Dirac en x_o et y_o respectivement. Alors posons

$$d'(\delta_{x_o}, \delta_{y_o}) = \sup\{ |f(x_o) - f(y_o)| \ / \ \|f\|_{BL} \leq 1 \text{ et } \|f\|_\infty = \frac{1}{2} |f(x_o) - f(y_o)| \} .$$

On a : $d_{BL}(\delta_{x_o}, \delta_{y_o}) = d'(x_o, y_o)$.

<u>Preuve</u> : on voit facilement qu'il existe de telles fonctions et que $d_{BL} \geq d'$.

Inversement : soit f telle que $\|f\|_{BL} \leq 1$.
On peut supposer que f est ≥ 0 pour x_o ou pour y_o (car sinon $-f$ est ≥ 0 et vérifie $\|f\|_{BL} \leq 1$).

Supposons donc $f(y_o) > 0$ et $f(y_o) > f(x_o)$.

Posons $g_1(x) = f(y_o) \wedge f(x)$.

Alors $\|g_1\|_L \leq \max(\|f(y_o)\|_L , \|f\|_L) = \|f\|_L$.

Puis $g_2(x) = g_1(x) - \frac{1}{2} (f(y_o) + f(x_o))$. Enfin soit

$g(x) = g_2(x) \vee - \frac{1}{2} (f(y_o) - f(x_o))$.

Alors $\|g\|_L \leq \|f\|_L$. Donc $\|g\|_{BL} \leq 1$, $\|g\|_\infty = \frac{1}{2} (f(y_o) - f(x_o))$ et $g(y_o) - g(x_o) = f(y_o) - f(x_o)$. Donc $d_{BL} \leq d'$ d'où le résultat.

Soit donc f telle que $\|f\|_{BL} \leq 1$ et $\|f\|_{\infty} = \frac{1}{2} |f(x_o) - f(y_o)|$.

Posons $z = |f(x_o) - f(y_o)|$. Alors $\frac{z}{d(x_o, y_o)} \leq \|f\|_L \leq 1 - \frac{z}{2}$. Dans notre cas

$z \leq \frac{1}{2} (1 - \frac{z}{2})$ soit $z \leq \frac{2}{1 + 2n}$.

Donc $d_{BL}(\delta x_o, \delta y_o) \leq \frac{2}{1+2n}$. D'où

$$d_{BL}(\delta x_o, \delta y_o) = \frac{2}{1+2n} .$$

On a également $\pi(\delta x_o, \delta y_o) = d(x_o, y_o) \wedge 1$ quand $n \to 0$ $d_{BL}(\delta x_o, \delta y_o) \to$ alors que $\pi(\delta x_o, \delta y_o) = 1$.

Si $n = 1$, $\pi(\delta x_o, \delta y_o) = 1$ et $d_{BL}(\delta x_o, \delta y_o) = \frac{2}{3}$.

Les inégalités 9 sont donc optimales dans le cas général.

I.3.3. Calcul de la distance de Fortet-Mourier.

Nous allons maintenant proposer une méthode de calcul approché de $\|P-Q\|_L^*$ dans le cas général. Nous testerons cette méthode dans le cas $S = \mathbb{R}$ et traiterons un exemple dans $S = \mathbb{R}^2$.
Soient donc P et Q deux éléments de $\mathcal{P}_1(S)$.

Méthode de Calcul. On va construire deux suites (P_n) et (Q_n) à support fini telles que $\|P_n - P\|_L^* \to 0$ et $\|Q_n - Q\|_L^* \to 0$.
Puis on va calculer $\|P_n - Q_n\|_L^*$ et approcher $\|P-Q\|_L^*$ par cette valeur.

Nous allons tout d'abord énoncer un lemme qui permet à la fois de construire les suites (P_n) et (Q_n) et d'obtenir pour chaque n, une majoration de $\| P_n - P \|_L^*$ (resp. $\| Q_n - Q \|_L^*$).

11. <u>Lemme</u> : <u>Soient</u> φ <u>une application mesurable de</u> S <u>dans</u> S, ε <u>et</u> δ <u>deux réels positifs</u>. <u>Appelons</u> P_φ <u>la mesure image de</u> P <u>par</u> φ <u>et</u> $H = \{x \in S \,/\, d(x, \varphi(x)) < \varepsilon\}$.

<u>On suppose que</u> $\int_{S-H} d(x, \varphi(x))\, d\, P(x) < \delta$.

<u>Alors</u> $\| P - P_\varphi \|_L^* < \varepsilon + \delta$.

<u>Preuve</u> : $\| P - P_\varphi \|_L^* = \sup_{f \in \mathcal{F}} \left| \int f\, d(P - P_\varphi) \right| = \sup_{f \in \mathcal{F}} \int f\, d(P - P_\varphi)$

car si $f \in \mathcal{F}$ alors $-f \in \mathcal{F}$

$$= \sup_{f \in \mathcal{F}} \left(\int f\, dP - \int f\, d P_\varphi \right) = \sup_{f \in \mathcal{F}} \left(\int f\, dP - \int f \circ \varphi\, dP \right)$$

$$= \sup_{f \in \mathcal{F}} \int [\, f(x) - f(\varphi(x)\,]\, dP \leq \int d(x, \varphi(x))\, dP$$

$$= \int_H d(x, \varphi(x))\, dP + \int_{S-H} d(x, \varphi(x))\, dP \leq \varepsilon + \delta .$$

Fixons maintenant $n \in \mathbb{N}^*$. On cherche un compact K_n et un point x_{m+1} tels que $\int_{S-K_n} d(x, x_{m+1})\, d(P+Q) < 1/2n$.

(Notons qu'on pourra utilement choisir deux ou plusieurs points jouant le rôle de x_{m+1}).

Construisons alors une partition $(A_j)_{j=1,\dots,m}$ de K_n telle que $\forall j \;\; \exists\, x_j \in A_j$ avec $\forall x \in A_j \;\; d(x, x_j) < 1/4\, n$.

Définissons l'application $\varphi_n : S \longrightarrow S$ par

$$\varphi_n(x) = \begin{cases} x_j & \text{si} \quad x \in A_j \\ x_{m+1} & \text{si} \quad x \in S - K_n . \end{cases} \tag{12}$$

Appelons P_n (resp. Q_n) la mesure image de P (resp. Q) par φ_n .

13. <u>Corollaire</u> : <u>on a</u> $\| P - P_n \|_L^* + \| Q - Q_n \|_L^* < 1/n$.

Utilisons maintenant la décomposition de Hahn-Jordan

$$P_n = P_n \wedge Q_n + (P_n - Q_n)^+$$

$$Q_n = P_n \wedge Q_n + (P_n - Q_n)^- \quad .$$

Posons $(P_n - Q_n)^+ = P_{1n}$ et $(P_n - Q_n)^- = Q_{1n}$.

Les mesures positives P_{1n} et Q_{1n} sont orthogonales et l'on a :

$\forall f \in \mathcal{F} \quad \int f \, d(P_n - Q_n) = \int f \, d(P_{1n} - Q_{1n})$ ce qui conduit à

$$\| P_n - Q_n \|_L^* = \| P_{1n} - Q_{1n} \|_L^* \quad .$$

Introduisons alors quelques notations :

pour $j = 1, \ldots, m+1$ posons $P_{1n}(x_j) = p_{1j}$ et $Q_{1n}(x_j) = q_{1j}$.

Pour i, j éléments de $\{1, \ldots, m+1\}$ on pose $d_{ij} = d(x_i, x_j)$.

Si $f \in \mathcal{F}$ on note pour $i = 1, \ldots, m+1$, $f(x_i) = f_i$.

Puis l'on pose $I = \{ i \,/\, p_{1i} > 0 \}$

$J = \{ j \,/\, q_{1j} > 0 \}$.

On obtient alors :

$$\int f \, d(P_{1n} - Q_{1n}) = \sum_{i \in I} f_i \, p_{1i} - \sum_{j \in J} f_j \, q_{1j} \quad .$$

Pour pouvoir progresser dans le calcul de $\| P_{1n} - Q_{1n} \|_L^*$ nous allons démontrer quelques lemmes.

Soient I et J deux sous-ensembles de $\{1, \ldots, m+1\}$ tels que $I \cup J = \{1, \ldots, m+1\}$ et $I \cap J = \emptyset$.

Un élément g de $\mathbb{R}^{m+1}$ sera noté $g = (g_{1i},\ i \in I\ ;\ g_{2j}\ ,\ j \in J)$.

Soient $(r_{1i})_{i \in I}$ $(r_{2j})_{j \in J}$ et $(d_{ij})_{(i,j) \in \{1,\dots,m+1\} \times \{1,\dots,m+1\}}$ des suites de réels ≥ 0 donnés. On définit sur $\mathbb{R}^{m+1}$ la forme linéaire $W(g)$ par :

$$W(g) = \sum_{i \in I} g_{1i}\, r_{1i} - \sum_{j \in J} g_{2j}\, r_{2j} \quad \text{sous les contraintes}$$

$$(14) \qquad \left.\begin{array}{l} \forall i \in I \\ \forall j \in J \end{array}\right\} \quad g_{1i} - g_{2j} \leq d_{ij} \ .$$

15. <u>Lemme</u> : Posons :

$$(16) \qquad \left\{\begin{array}{ll} \forall i \in I & f_{1i} = \min\limits_{k \in J} (g_{2k} + d_{ik}) \\ \forall j \in J & f_{2j} = \max\ (f_{1h} - d_{hj}) \ . \end{array}\right.$$

Alors la suite $f = (f_{1i},\ i \in I\ ;\ f_{2j}\ ,\ j \in J) \in \mathbb{R}^{n+1}$ vérifie les contraintes (14)et l'on a $W(f) \geq W(g)$.

<u>Preuve</u> : pour $i \in I$ et $j \in J$ donnés, on a :

$$f_{1i} - f_{2j} = \min_{k \in J} (g_{2k} + d_{ik}) - \max_{h \in I} (f_{1h} - d_{hj})$$

$$= \min_{h \in J} (g_{2k} + d_{ik}) - \max_{h \in I} \left(\min_{k \in J} (g_{2k} + d_{hk}) - d_{hj} \right)$$

$$\leq \min_{k \in J} (g_{2k} + d_{ik}) - \min_{k} (g_{2k} + d_{ik}) - d_{ij} = d_{ij} \ .$$

Les contraintes (14) sont donc vérifiées. D'autre part

$$\forall j \in J\ ,\quad g_{2j} + d_{ij} \geq g_{1i} \Longrightarrow f_{1i} \geq g_{1i}$$

et $\forall j \in J \quad f_{2j} = \max\limits_{h \in I} \left(\min\limits_{k \in J} (g_{2k} + d_{hk}) - d_{hj} \right) \leq \max\limits_{h \in I} (g_{2j} + d_{hj} - d_{hj}) = g_{2j}$

entraînent $W(f) \geq W(g)$.

17. <u>Lemme</u> : <u>Reprenons les hypothèses du lemme 15</u>. <u>Si de plus la suite</u> (d_{ij}) <u>vérifie l'inégalité triangulaire alors</u>

$\forall i \in I$ et $j \in J$ $\quad |f_{1i} - f_{2j}| \leq \max (d_{ij} , d_{ji})$

$\forall i_1, i_2$ éléments de I $\quad |f_{1i_1} - f_{1i_2}| \leq \max (d_{i_1 i_2} , d_{i_2 i_1})$

$\forall j_1, j_2$ éléments de J $\quad |f_{2j_1} - f_{2j_2}| \leq \max (d_{j_1 j_2} , d_{j_2 j_1})$.

<u>Preuve</u> : soient $i \in I$ et $j \in J$ donnés. Supposons que $\min\limits_{k \in J} (g_{2k} + d_{ik}) = g_{2k_o} + d_{ik_o}$

alors $f_{1i} - f_{2j} = g_{2k_o} + d_{ik_o} - \max\limits_{h \in I} (g_{2k_o} + d_{hk_o} - d_{hj})$

$= d_{ik_o} - \max\limits_{h \in I} (d_{hk_o} - d_{hj}) \geq d_{ik_o} - d_{jk_o} \geq -d_{ij}$.

On déduit alors la première inégalité à l'aide des résultats du lemme 7.

Soient maintenant i_1 et i_2 deux éléments de I

$f_{1i_1} - f_{2i_2} = \min\limits_{k \in J} (g_{2k} + d_{i_1 k}) - \min\limits_{k \in J} (g_{2k} + d_{i_2 k})$.

Supposons que $\min\limits_{k \in J} (g_{2k} + d_{i_2 k}) = g_{2k_o} + d_{i_2 k_o}$. Alors

$f_{1i_1} - f_{2i_2} \leq g_{2k_o} + d_{i_1 k_o} - (g_{2k_o} + D_{i_2 k_o}) = d_{i_1 k_o} - d_{i_2 k_o} \leq d_{i_1 i_2}$.

Inversement, supposons que $\min\limits_{k \in J} (g_{2k} + d_{i_1 k}) = g_{2k_1} + d_{i_1 k_1}$

alors $f_{1i_1} - f_{1i_2} \geq g_{2k_1} + d_{i_1k_1} - (g_{2k_1} + d_{i_2k_1}) = d_{i_1k_1} - d_{i_2k_1} \geq - d_{i_2i_1}$.

On en conclut la deuxième inégalité.

De la même façon on démontre la troisième inégalité.

18. Corollaire : Supposons que $\forall i$ et j élément 0 de $\{1,\ldots,m+1\}$, $d_{ij} = d(x_i, x_j)$ où d est la distance sur S . Les problèmes ① et ② ci-dessous sont équivalents :

problème 1 : maximiser $W(f) = \sum_{i\in I} f_{1i}\, r_{1i} - \sum_{j\in J} f_{2j}\, r_{2j}$
f

sous les contraintes $\forall i \in I$, $\forall j \in J$ $\quad |f_{1i} - f_{2j}| \leq d_{ij}$

$\forall i_1, i_2$ éléments de I $\quad |f_{1i_1} - f_{1i_2}| \leq d_{i_1i_2}$

$\forall j_1, j_2$ éléments de J $\quad |f_{2j_1} - f_{2j_2}| \leq d_{j_1j_2}$.

problème 2 : maximiser $W(f)$ sous les contraintes :

$$\forall i \in I\ ,\ \forall j \in J\ ,\ f_{1i} - f_{2j} \leq d_{ij}\ .$$

Preuve : de façon claire $\max_{①} \leq \max_{②}$ car une suite f qui vérifie les contraintes du problème ① vérifie celles du problème ② . Inversement, à partir d'une suite g vérifiant les contraintes du problème 2 on peut construire d'après les lemmes 17 et 18 et compte tenu des hypothèses sur les (d_{ij}) une suite f qui vérifie les contraintes du problème ① et telle que $W(f) \geq W(g)$.

On voit alors qu'en imposant $\sum_{i\in I} r_{1i} = \sum_{j\in J} r_{2j}$ le calcul de $\| P_{1n} - Q_{1n} \|_L^*$ se ramène aux problèmes précédents.

Remarque : A la place de la construction (16) on aurait pu faire la cons-

truction (19) ci-dessous :

$$\begin{cases} \forall j \in J & f_{2j} = \max\limits_{i \in I} (g_{1i} - d_{ij}) \\ \forall i \in I & f_{1i} = \min\limits_{j \in J} (f_{2j} + d_{ij}) \end{cases} .$$

On vérifie facilement que la suite ainsi définie vérifie les hypothèses des lemmes 15 et 18. Mais l'une des deux suites est plus performante que l'autre, en ce sens qu'elles donnent en général des valeurs de $W(f)$ différentes comme on le voit facilement sur l'exemple 20.

20. <u>Exemple</u>. $S = \mathbb{R}^2$ d est la distance euclidienne

$x_1 = (0,0)$ $x_2 = (1,0)$ $x_3 = (1,1)$ $x_4 = (2,1)$

$I = \{1,2\}$ $J = \{3,4\}$

$r_{11} = 1/4 = r_{12}$ $r_{23} = 1/6$ $r_{24} = 2/6$

$g_{11} = 1$ $g_{12} = 2$ $g_{23} = 3$ $g_{24} = 6$

$d_{12} = 1$, $d_{13} = \sqrt{2}$, $d_{14} = \sqrt{5}$, $d_{23} = 1$, $d_{24} = \sqrt{2}$, $d_{34} = 1$

appelons f^2 la suite obtenue sous la construction (16) et f^1 la suite obtenue sous la construction (19)

$f^1_{11} = 1 + \sqrt{2}$ $f^1_{12} = 2$ $f^1_{23} = 1$ $f^1_{24} = 2-\sqrt{2}$

$f^2_{11} = 3 + \sqrt{2}$ $f^2_{12} = 4$ $f^2_{23} = 3$ $f^2_{24} = 4-\sqrt{2}$

$$W(f^1) = \frac{7\sqrt{2}-1}{12} \qquad W(f^2) = \frac{7+5\sqrt{2}}{12} .$$

Revenons au calcul de $\| P_{1n} - Q_{1n} \|^*_L$. Nous prolongeons P_{1n} et Q_{1n} sur $\{x_1, \ldots, x_{m+1}\}$ en posant

$$\forall j \in J \qquad P_{1n}(x_j) = p_{1j} = 0$$

$$\forall i \in I \qquad Q_{1n}(x_i) = q_{1i} = 0 \ .$$

On a alors la succession de problèmes équivalents suivants :
(les résultats sur la dualité pourront être vus dans Simonnard [13]).
La chaîne des équivalences est la suivante :

$$① \Longleftrightarrow ② \Longleftrightarrow ③ \Longleftrightarrow ④ \Longleftrightarrow ④' \Longleftrightarrow ③' \ .$$

On en déduit que ②⟺②' en tenant compte des valeurs nulles qui interviennent et finalement ①⟺②' .

problème primal :	problème dual
① maximiser $\sum_{i \in I} f_i \, p_{1i} - \sum_{j \in J} f_j \, q_{1j}$ sous les contraintes : $\forall (i,j) \in \{1,\dots,m+1\} \times \{1,\dots,m+1\}$ $\lvert f_i - f_j \rvert \leq d_{ij}$	①' = ②' compte tenu du corollaire 11
② maximiser $\sum_{i \in I} f_{1i} \, p_{1i} - \sum_{j \in J} f_{2j} \, q_{1j}$ sous les contraintes : $\forall i \in I \ , \ \forall j \in J \quad f_{1i} - f_{2j} \leq d_{ij}$	②' minimiser $\sum_{j \in I} x_{ij} \, d_{ij}$ sous les contraintes : $x_{ij} \geq 0 \quad \sum_{i \in I} x_{ij} = q_{1j}$ $\sum_{j \in J} x_{ij} = p_{1i}$ ②' est obtenu à partir de ③' en tenant compte des valeurs des (p_{1i}) et (q_{1j})
③ maximiser $\sum_{k=1} (f_{1k} \, p_{1k} - f_{2k} \, q_{1k})$ sous les contraintes : $\forall i$ et j éléments de $\{1,\dots,m+1\}$ $f_{1i} - f_{2j} \leq d_{ij}$	③' = ④'

④ maximiser $\sum_{k=1}^{m+1} (f_{1k}\ p_{1k} + f_{2k}\ q_{1k})$

sous les contraintes :

$\forall i,j$ éléments de $\{1,\dots,m+1\}$

$f_{1i} + f_{2j} \leq d_{ij}$

④' minimiser $\sum_{i,j \in \{1,\dots,m+1\}} x_{ij}\ d_{ij}$

sous les contraintes :

$x_{ij} \geq 0$

$\sum_{i=1}^{m+1} x_{ij} = q_{1j}$

$\sum_{j=1}^{m+1} x_{ij} = p_{1i}$

On est donc ramené à résoudre le problème ②' qui est un problème de transport pour lequel il existe plusieurs algorithmes.
Pour finir notons que :

$$\|P_n - Q_n\|_L^* - \|P - P_n\|_L^* - \|Q - Q_n\|_L^* \leq \|P - Q\|_L^* \leq \|P - P_n\|_L^* + \|P_n - Q_n\|_L^* + \|Q_n - Q\|_L^*$$

d'où l'on tire compte tenu du corollaire 6

$$\left| \ \|P - Q\|_L^* - \|P_n - Q_n\|_L^* \ \right| < 1/n \ .$$

21. <u>Remarques</u> : Supposons par exemple que l'on ait numéroté les (x_j) de telle sorte que $I = \{1,\dots,n_o\}$ et $J = \{n_o+1,\dots,m+1\}$.
La matrice des x_{ij} va se présenter sous la forme

$$(x_{ij}) = \begin{array}{c} \\ 1 \\ \\ n_o \\ \\ m+1 \end{array} \begin{array}{c} \begin{array}{cc} n_o & n_o+1 \quad\quad m+1 \end{array} \\ \left(\begin{array}{c|c} 0 & A \\ \hline 0 & 0 \end{array}\right) \end{array}$$

A sera une matrice $n_o \times (m+1-n_o)$.

Pour les $n_o \times (m+1-n_o)$ inconnues du problème dual, on dispose de

$(m+1-n_o) + n_o = m+1$ équations. Dans certains cas il y a unicité de la solution comme on peut le voir sur l'exemple suivant :

22. <u>Exemple</u> : sur $\mathbb{R}$ muni de la distance euclidienne prenons

$$x_1 = 1 \qquad x_2 = 2 \qquad x_3 = 4 \qquad x_4 = 4$$

on obtient pour $(d_{ij}) = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix}$

soient $p_{11} = 3/16 \qquad p_{12} = 2/16 \qquad p_{13} = 1/16 \qquad p_{14} = 0$

$q_{11} = 0 \qquad q_{12} = 0 \qquad q_{13} = 0 \qquad q_{14} = 6/16$.

Alors il y a unicité de la solution (x_{ij}) donnée par :

$$(x_{ij}) = \left(\begin{array}{c|c} & 3/16 \\ 0 & 2/16 \\ & 1/16 \\ \hline 0 & 0 \end{array}\right) .$$

On en déduit que $\min\limits_{\substack{i=1,2\,3 \\ j=4}} x_{ij}\, d_{ij} = 14/16$ sous les contraintes du problème (2').

On peut alors comparer la rapidité de ce calcul avec celle du résultat de Wallender [15] dans le cas de quatre points $\{x_1, x_2, x_3, x_4\}$ sur lesquels on a deux probabilités P et Q de valeurs respectives $p_1, \ldots, p_4$; $q_1, \ldots, q_4$ et sous les hypothèses supplémentaires :

(23) $\quad d(x_1, x_3) + d(x_2, x_4) \geq \max\,(\, d(x_1, x_2) + d(x_3, x_4)\,;\, d(x_1, x_4) + d(x_2, x_3)\,)$

$$2\,\|P-Q\|_L^* \doteq d_{12}\,[\,|p_1-q_1| + |p_2-q_2| - |p_1+p_2-q_1-q_2|\,]$$
$$+\, d_{14}\,[\,|p_1-q_1| + |p_4-q_4| - |p_1+p_4-q_1-q_4|\,]$$
$$+\, d_{23}\,[\,|p_2-q_2| + |p_3-q_3| - |p_2+p_3-q_2-q_3|\,]$$

$$+ d_{34} \left[|p_3 - q_3| + |p_4 - q_4| - |p_3 + p_4 - q_3 - q_4| \right]$$

$$+ d_{13} \left[|p_1 + p_2 - q_1 - q_2| + |p_1 + p_4 - q_1 - q_4| - |p_2 - q_2| - |p_4 - q_4| \right]$$

$$+ d_{24} \left[|p_1 + p_2 - q_1 - q_2| + |p_1 + p_4 - q_1 - q_4| - |p_1 - q_1| - |p_3 - q_3| \right]$$

24. <u>Etude sur</u> $\mathbb{R}$.

Sur $\mathbb{R}$ muni de la distance euclidienne considérons les probabilités P et Q définies par les densités respectives f et g

$$f(x) = e^{-x} \mathbb{1}_{R+}(x)$$

$$g(x) = \frac{1}{2} e^{-x/2} \mathbb{1}_{R+}(x)$$

n étant donné, on va tout d'abord chercher x_o tel que

$\delta_2 = \int_{x_o}^{+\infty} (x - x_o) \frac{1}{2} e^{-x/2} dx < \frac{\alpha}{4n}$ où α est un coefficient proche de 2 . En posant $u = (x - x_o)/2$ on obtient

$$\delta_2 = 2 \int_0^{+\infty} u \; e^{-(u + x_o/2)} du < \frac{\alpha}{4n} \quad \text{soit} \quad 2 e^{-x_o/2} \Gamma(2) < \frac{\alpha}{4n}$$

ou encore :

$$x_o > 2 \operatorname{Log} \frac{8n}{\alpha} .$$

On a également $\delta_1 = \int_{x_o}^{+\infty} (x - x_o) e^{-x} dx = e^{-x_o}$. On vérifie alors que

$e^{-x_o} < \frac{2 - \alpha}{4n}$ de telle sorte que $\int_{x_o}^{+\infty} (x - x_o) (e^{-x} + \frac{1}{2} e^{-x/2}) dx < \frac{1}{2n}$.

On divise alors $[0, x_o]$ en intervalles de longueur $1/2n$.
Le calcul de la charge de ces intervalles, à savoir :

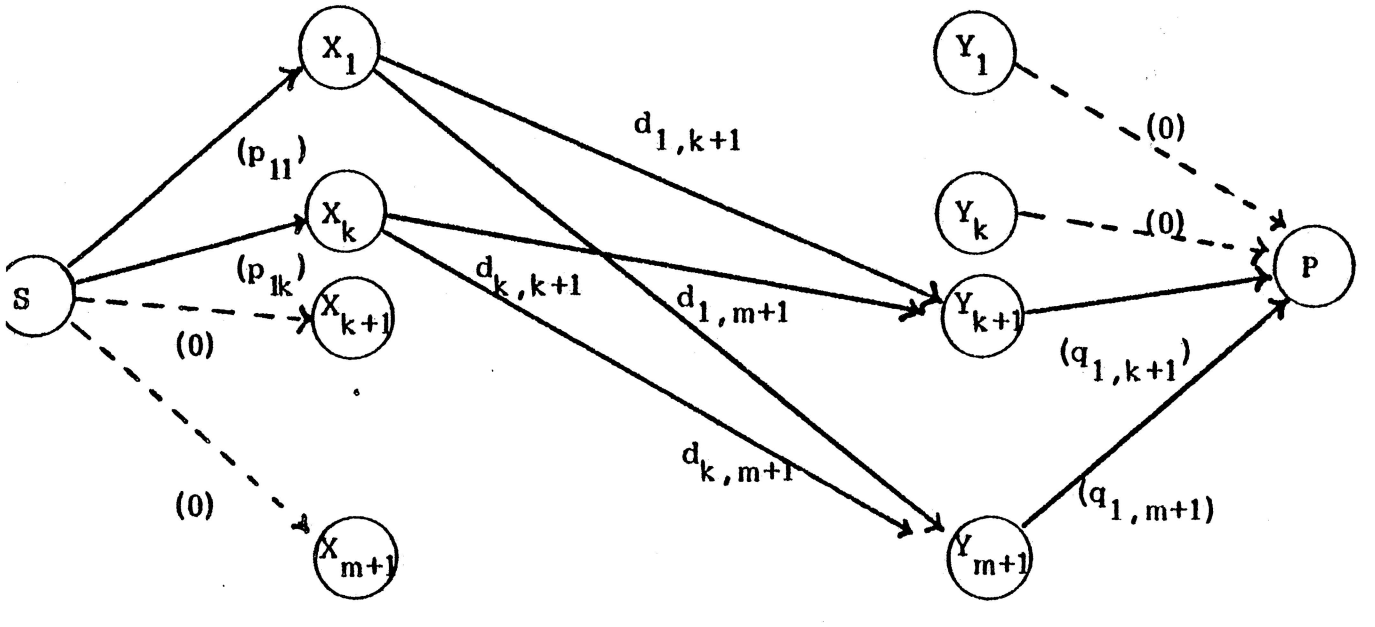
$$\int_{(i-1)/2n}^{i/2n} e^{-x}\,dx \quad \text{et} \quad \int_{i-1/2n}^{i/2n} \frac{1}{2}\, e^{-x/2}\,dx = e^{\frac{i-1}{4n}} - e^{\frac{i}{4n}}$$

est effectué en double précision. Les charges associées à P_{in} nous sont alors fournies par :

$$\max\left(\int_{i-1/2n}^{1/2n} e^{-x}\,dx - \int_{\frac{i-1}{2n}}^{\frac{i}{2n}} \frac{1}{2}\, e^{-x/2}\,dx \;;\; 0\right)$$

et celles de Q_{in} par : $\max\left(0 \;;\; \int_{\frac{i-1}{2n}}^{\frac{i}{2n}} \frac{1}{2}\, e^{-x/2}\,dx - \int_{\frac{i-1}{2n}}^{\frac{i}{2n}} e^{-x}\,dx\right)$

On multiplie ces charges par une puissance de 2 (2^{26} par exemple). On ne conserve de ce produit que la partie entière. Les erreurs d'arrondis sont reportées sur la charge du dernier intervalle et sont de toute façon négligeables pour la précision de nos calculs. Les distances entre le milieu x_i de l'intervalle $I_i = [\frac{i-1}{2n}, \frac{i}{2n}[$ et le milieu x_j de l'intervalle I_j sont multipliées par $2n$, de telle sorte que le programme tourne en nombres entiers. Posons $m = [2n.x_o]$. Pour expliciter les modalités du calcul construisons le graphe suivant :



CALCUL DE LA DISTANCE DE FORTET - M - SUR $\mathbb{R}$

avec P de densité : $f(x) = e^{-x} \mathbb{1}_{\mathbb{R}^+}(x)$ et Q de densité $g(x) = \frac{1}{2} e^{-x/2} \mathbb{1}_{\mathbb{R}^+}(x)$

abscisse du Point d'intersection des 2 courbes représentatives des densités : 2 Log2 .

n	Extrémité des intervalles: amax	Majoration de $\delta_1 + \delta_2$	Nombre de sommets X_i	Nombre de sommets Y_j	Taille de la matrice (d_{ij})	Nombre d'itérations	Temps de calcul (en secondes	coût minimal
1	10		3	17	51	1	0,278	0,97449
10	8,8	0,0247	28	148	4 144	1		0,975
20	8,9	0,0235	55	301	16 555	1		0,9766
1	12		3	21	63	1	0,641	0,98406
20	12	0,0050	55	425	23 375	1	(#4)	0,99499

En utilisant la formule de Fortet-Mourier on a :

$$||P - Q||_L^* = \int_0^{+\infty} |1 - e^{-x} - 1 + e^{-x/2}| dx = \int_0^{+\infty} (e^{-x/2} - e^{-x}) dx$$

$$= 2 e^{-x/2} \Big|_0^{+\infty} + e^{-x} \Big|_0^{+\infty} = 1 .$$

On veut faire passer le flot sortant de S et qui vaut $\sum_{\ell=1}^{k} p_{1\ell}$ vers P où peut arriver $\sum_{k+1}^{m+1} q_{1\ell}$ (ces deux sommes étant égales) pour un coût minimal. Le coût de transport sur chaque arc (X_i, X_j) est égal à la distance $d_{ij} = d(X_i, Y_j)$. La capacité maximale d'un arc est entre parenthèse. La capacité maximale de l'arc (S, X_i) exprime que $\sum_j x_{ij} \leq p_{1i}$. De même la capacité maximale de l'arc (Y_j, P) exprime que $\sum_i x_{ij} \leq q_{ij}$. Pour obtenir le flot de coût minimal, on utilise le programme de la bibliothèque Nag pour le problème de Hitchcock avec un ordinateur C.I.I.-HB 68. Les résultats des divers calculs sont consignés au tableau ci-contre.

De ces calculs, il ressort que la source d'erreur la plus importante est l'erreur de troncature pour Q, qui semble déterminer la précision du calcul, alors que la modification du pas (longueur des intervalles de charges) n'entraîne qu'une très légère modification du résultat.
La majoration des erreurs pour $n = 20$ et amax = 12 est de 0,055, alors que l'erreur absolue est de l'ordre de 5×10^3.

Signalons dès à présent l'intérêt de cette méthode sur R : d'une part elle peut s'utiliser avec une distance quelconque, d'autre part, dans le cas de deux lois empiriques sur un grand nombre de points, l'utilisation de la formule de Wallender [15] risque de s'avérer particulièrement fastidieux, alors que cette méthode donne la réponse exacte de façon quasi-immédiate.

25. Etude sur R^2.

Nous allons maintenant travailler sur R^2.
On munit R^2 de la distance associée à la norme :

$$\text{si } \underline{x} \in R^2, \quad \|\underline{x}\| = \max_{i=1,2} (|x_i|).$$

Nous prendrons pour P la loi normale centrée de matrice de variances-

covariance : l'identité, et pour Q la loi normale centrée de matrice de variances-covariances : $\begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$.

Commençons par énoncer quelques lemmes qui vont nous permettre de réduire le problème :

26. <u>Lemme</u> : <u>la norme ci-dessus est fonction croissante de la valeur absolue de chacune des composantes des vecteurs</u> $\in \mathbf{R}^2$.

<u>Preuve</u> : étant donnés x et y deux vecteurs de $\mathbf{R}^2$

$$\underline{x} = (x_1, x_2) \quad \text{et} \quad \underline{y} = (y_1, y_2) .$$

Supposons $|y_1| \geq |x_1|$ et $|y_2| \geq |x_2|$ alors

$$\max_{i=1,2} (|y_i|) \geq \max_{i=1,2} |x_i| - \text{i.e)} . \|\underline{y}\| \geq \|\underline{x}\| .$$

27. <u>Lemme</u> : <u>Soient</u> P <u>et</u> Q <u>deux mesures sur</u> $\mathbf{R}^2$, <u>symétriques par rapport à</u> Ox <u>et</u> Oy . <u>Appelons alors</u> $\mathcal{F}_s$ <u>le sous-ensemble de</u> $\mathcal{F}$ <u>formé des fonctions symétriques par rapport à</u> Ox et Oy . <u>On a</u>

$$\sup_{f \in \mathcal{F}} \int_{\mathbf{R}^2} f \, d(P-Q) = \sup_{f \in \mathcal{F}_s} \int_{\mathbf{R}^2} f \, d(P-Q) .$$

<u>Preuve</u> : il est clair que $\sup_{f \in \mathcal{F}} \int f \, d(P-Q) \geq \sup_{f \in \mathcal{F}_o} \int f \, d(P-Q)$. Inversement, supposons qu' $\exists$ f telle que $\int f \, d(P-Q) > \sup_{f \in \mathcal{F}_o} \int f \, d(P-Q)$.

Appelons f^i la restriction de f ou quadrant noté $qu(i)$ tel que $\int_{qu(i)} f^i \, d(P-Q) \geq \int_{qu(j)} f \, d(P-Q)$ pour $j = 1,2,3,4$.

Construisons à partir de f^i la fonction $\tilde{f}^i \in \mathcal{F}_s$ définie sur $\mathbf{R}^2$ telle que

$$\tilde{f}^1(\varepsilon_1 x_1, \varepsilon_2 x_2) = f^1(x_1, x_2) \quad \text{où}$$

$(x_1, x_2) \in qu(i)$ et $\varepsilon_1 = \pm 1$ et $\varepsilon_2 = \pm 1$. On a :

$$|f(\varepsilon_1 x_1, \varepsilon_2 x_2) - f(\varepsilon_3 y_1, \varepsilon_4 y_2)| = |f(x_1, x_2) - f(y_1, y_2)|$$

$$\leq d[(x_1, x_2), (y_1, y_2)] \leq d((\varepsilon_1 x_1, \varepsilon_2 x_2), (\varepsilon_3 y_1, \varepsilon_4 y_2))$$

d'après le lemme 26.

Pour terminer, remarquons qu'alors

$$\int \tilde{f}^1 \, d(P-Q) \geq \int f \, d(P-Q)$$

en contradiction avec l'hypothèse. Donc

$$\sup_{f \in \mathcal{F}} \int f \, d(P-Q) \geq \operatorname*{Sup}_{f \in \mathcal{F}} \int f \, d(P-Q)$$

d'où le résultat.

28. <u>Corollaire</u> : <u>appelons</u> P^1 <u>et</u> Q^1 <u>les restrictions respectives au premier quadrant de</u> P <u>et</u> Q, <u>mesures sur</u> R^2 <u>répondant aux hypothèses du lemme 23</u>. <u>Alors</u>

$$\|P-Q\|_L^* = 4\|P^1 - Q^1\|_L^* .$$

<u>Calcul du rayon des cercles d'intersection</u> :

la densité $f(x_1, x_2)$ de P vaut $\dfrac{1}{2\pi} e^{-\frac{x_1^2 + x_2^2}{2}}$

la densité $g(x_1, x_2)$ de Q vaut $\dfrac{1}{2\pi \cdot \sigma^2} e^{-\frac{x_1^2 + x_2^2}{2\sigma^2}}$

en posant $x_2 = 0$ on obtient l'abscisse des points d'intersection des deux surfaces dans le plan $x_2 = 0$ et donc le rayon cherché .

$$e^{-\frac{x_1^2}{2}} = \frac{1}{\sigma^2} \; e^{-\frac{x_1^2}{2\sigma^2}} \quad \text{ou encore} \quad \sigma^2 = e^{-\frac{x_1^2}{2}} \left(1 - \frac{1}{2\sigma^2}\right)$$

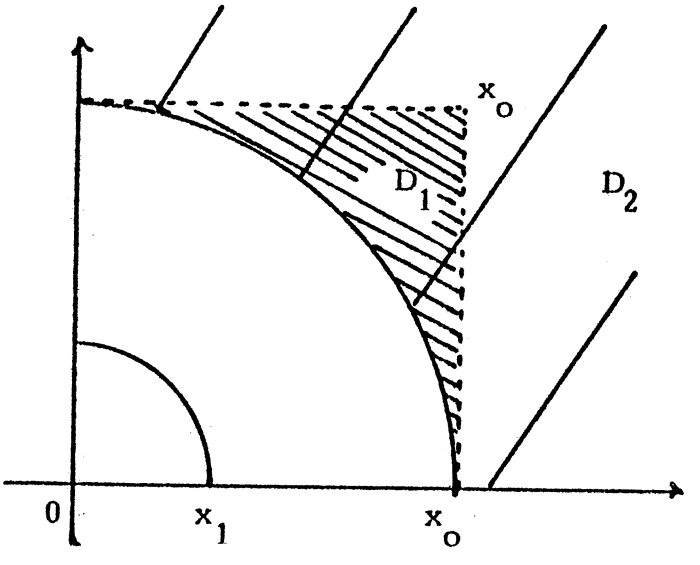
ce qui conduit à

$$2 \log \sigma = \frac{x_1^2}{2} \left(1 - \frac{1}{\sigma^2}\right)$$

ou

$$x_1 = \pm\, 2\sigma \sqrt{\frac{\operatorname{Log} \sigma}{\sigma^2 - 1}} \; .$$

On recherche alors une majoration convenable de l'erreur de troncature



notons $\underline{x_o} = (x_o, x_o)$

appelons D le domaine du plan défini par

$$D = \{ (y_1, y_2) \;/\; y_1 \geq 0 \; ; \; y_2 \geq 0 \; ; \; y_1^2 + y_2^2 > x_o^2 \} \; .$$

Si $\rho = \sqrt{y_1^2 + y_2^2}$ on vérifie que dans la zone hachurée en bleu D_1 , $d(\underline{x_o}, \underline{y}) \leq \rho$.

D'autre part soit $D_2 = D - D_1$ et $\underline{y} \in D_2$

$$d(\underline{x_o}, \underline{y}) = \max(|y_1 - x_o|, |y_2 - x_o|)$$

$$d^2(\underline{x_o}, \underline{y}) \leq (y_1 - x_o)^2 + (y_2 - x_o)^2 \leq y_1^2 + y_2^2 = \rho^2 \quad \text{car}$$

$y_1 \geq x_o$ et $y_2 \geq x_o$

donc $\forall \underline{y} \in D \quad d(\underline{x_o}, y) \leq \rho$.

Calculons alors $\delta_2 = \int_D d(\underline{x_o}, \underline{y})\, dQ$

$$= \iint_D d(x_o, y) \frac{1}{2\pi\sigma^2} e^{\frac{y_1^2 + y_2^2}{2\sigma^2}} dy_1 \, dy_2 .$$

Passons en coordonnées polaires. L'intégrale devient :

$$\int_0^{\pi/2} \frac{1}{2\pi} d\theta \int_{x_o}^{+\infty} d(\underline{x_o}, \underline{y}) \frac{\rho}{\sigma^2} e^{-\rho^2/2\sigma^2} d\rho$$

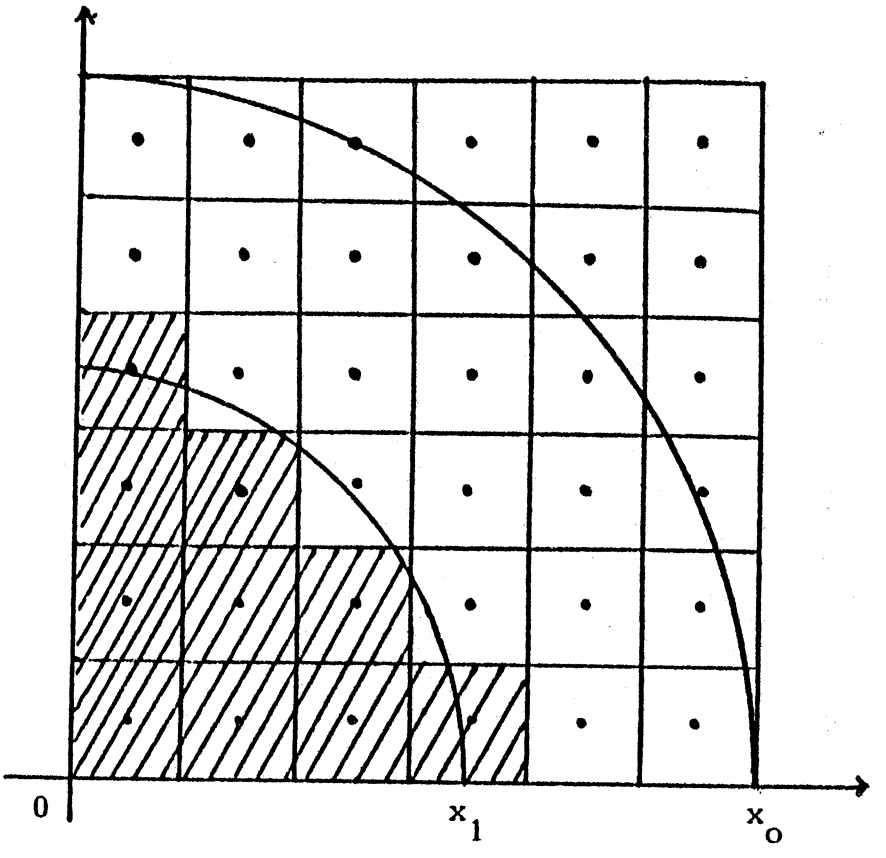
$$\leq \int_0^{\pi/2} \frac{1}{2\pi} d\theta \int_{x_o}^{+\infty} \rho^2/\sigma^2 \; e^{-\rho^2/2\sigma^2} d\rho$$

$$= \frac{\sigma}{4} \int_{x_o/\sigma}^{+\infty} u^2 e^{-\frac{u^2}{2}} du \quad \text{en ayant posé} \quad u = \frac{\rho}{\sigma} .$$

De même on trouve que $\delta_1 = \int_D d(\underline{x_o}, \underline{y})\, dP < \frac{1}{4} \int_{x_o}^{+\infty} u^2 e^{u^2/2} du$.

δ_1 est, d'ailleurs tout à fait négligeable devant δ_2 .

On construit alors le quadrillage suivant



On calcule alors successivement en double précision par la méthode de Gauss-Legendre

$$\int_{\frac{i-1}{invpas}}^{\frac{i}{invpas}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2\sigma^2} dx = \int_{\frac{i-1}{invpas}}^{\frac{i}{invpas}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

et $$\int_{\frac{i-1}{invpas}}^{\frac{i}{invpas}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx .$$

On en déduit alors la charge affectée à chacun des carrés du quadrillage par P_n et Q_n puis la charge affectée à chacun des carrés par P_{1n} et Q_{1n} . On multiplie cette charge par une puissance de 2:(2^{34}) par

exemple et l'on ne garde que la partie entière de ce produit. La largeur des carrés du quadrillage est égale à invpas

Compte tenu des résultats sur $\mathbb{R}$, on a cherché à obtenir une bonne majoration de l'erreur de troncature $\delta_1 + \delta_2$ plutôt qu'à prendre une très petite valeur pour la largeur de ces carrés. Les résultats sont consignés dans le tableau ci-contre.

Notons pour terminer, que cette méthode donne la valeur exacte de $\| P-Q \|_L^*$ pour toutes lois P et Q à support fini dans un espace métrique quelconque.

CALCUL DE LA DISTANCE DE FORTET-M-SUR $\mathbb{R}^2$

P de loi $\mathcal{N}\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ de densité $\frac{1}{2\pi} e^{-\frac{y_1^2+y_2^2}{2}}$

Q de loi $\mathcal{N}\left(0, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right)$ de densité $\frac{1}{2\pi\sigma^2} e^{-\frac{y_1^2+y_2^2}{2\sigma^2}}$

σ	$\gamma_{\delta 1}$	a_{max}	Majoration de $\delta_1+\delta_2$	Inv-pas	Nbre de points X_i	Nbre de points Y_j	Taille matrice (d_{ij})	Temps en min.	Nbre d'itérations	$\lVert P^1-Q^1 \rVert_L^*$	$\lVert P-Q \rVert_L^*$
3	2,2235	7	0,133	1	4	45	180	0,009	4	0,54871	2,19484
3	2,2235	7,33	0,106	3	37	447	16 539	2,09	183	0,55583	2,22332
3	2,2235	10	0,010	2	15	385	5 775	0,41	76	0,56374	2,2550
4	2,4320	14	0,008	2	19	765	14 535	1,78	149	0,84601	3,3840
5	2,5896	16	0,026	2	22	1 002	22 044	4,46	284	1,12731	4,509
6	2,7151	20	0,021	2	22	1 578	34 716	8,07	292	1,40969	5,63876

I.3.4. Un aspect du théorème de Kantorovitch et Rubinstein.

29. Introduction : Kantorovitch et Rubinstein ont démontré en 1958 le résultat suivant (d'après Veršik [15]).
Soit (S_K, d) un espace métrique compact. Appelons $\mathcal{L}$ l'espace des applications mesurables définies sur $([0,1], \mathcal{B}[0,1], dx)$ à valeurs dans (S_K, d).

Soit ℓ_ν le sous-ensemble de $\mathcal{L}$ dont les éléments ont ν pour mesure image. Notons $\|\cdot\|$ la norme de la variation totale. Soient ν_1 et ν_2 deux mesures sur S_K telles que $\|\nu_1\| = \|\nu_2\|$. Alors

$$(30) \qquad \|\nu_1 - \nu_2\|_L^* = \inf \{ \int_0^1 d(f(t), g(t))\, dt \ / \ f \in \ell_{\nu_1} \text{ et } g \in \ell_{\nu_2} \}.$$

Il s'agit en fait d'une variation du résultat de Kantorovitch et Rubinstein. Le travail original a été traduit par R.M. Dudley [8].

Soit maintenant (S,d) un espace Polonais et P et Q deux éléments de $\mathcal{P}_1(S)$.
En 1969 Wasserstein [17] a défini la distance entre P et Q suivante, notée W :

$$(31) \quad W(P,Q) = \inf \{ \int d(x,y)\, d\gamma \ / \text{ où } \gamma \in \mathcal{P}(S \times S) \quad \gamma \circ \pi_1^{-1} = P \quad \gamma \circ \pi_2^{-1} = Q \}$$

π_i $(i=1,2)$ étant la projection canonique sur la $i^{\text{ème}}$ composante.

En 1970, Dobrushin [4] établit le lien suivant : soit $P \in \mathcal{P}(S)$ et x_o donnés. Si l'on appelle $\mathcal{R}$ l'ensemble :

$$\mathcal{R}(P) = \{ Q \in \mathcal{P}(S) \ / \ W(P,Q) < +\infty \text{ et } \lim_{C \to +\infty} \int_{\{x / d(x_o, x) > C\}} d(x, x_o)\, dQ = 0 \}.$$

Alors la topologie induite par W sur $\mathcal{R}(P)$ est équivalente à la topologie de la convergence faible.

Wallender obtient en 1972, par une démonstration directe que sur $(\mathbb{R}, |.|)$

$$W(P,Q) = \int_{-\infty}^{+\infty} |F(x) - G(x)| \, dx \, . \tag{32}$$

En 1976, Dudley [7] propose une démonstration du théorème de Kantorovitch et Rubinstein pour (S,d) un espace polonais. Elle est complétée mais non publiée dans [9]. Une démonstration complète est finalement publiée en 1982 par de Acosta [1].

On se propose ici, de travailler sur l'espace probabilisé $[0,1]$ muni de la tribu des boréliens et de la mesure de Lebesgue notée dx. Nous considérons alors les variables aléatoires définies sur $([0,1], \mathcal{B}[0,1], dx)$ à valeurs dans l'espace polonais (S,d). Nous voulons montrer le résultat suivant :

Soient P et Q deux probabilités à support fini sur $(S, \mathcal{B})$.
Alors pour toute variable aléatoire X de $[0,1]$ dans S de loi P, il existe une variable aléatoire Y de $[0,1]$ dans S de loi Q telle que

$$\| P - Q \|_L^* = E\,(d(X,Y)) \, .$$

On étudie alors ce que devient ce résultat lorsque P et Q sont sans atomes.

33. <u>Construction de</u> X <u>et de</u> Y. Précisons, tout d'abord, les notations.

- S,d est un espace métrique séparable et complet. Nous appelons $\mathcal{B}$ sa tribu borélienne.

- X et Y des variables aléatoires définies sur $([0,1], \mathcal{B}[0,1], dx)$ à valeurs dans $(S, \mathcal{B})$.

- P et Q sont deux probabilités sur $(S, \mathcal{B}) \in \mathcal{P}_1(S)$.

- $L^1([0,1], \mathcal{B}[0,1], dx, S, d)$ est l'ensemble des classes d'équivalence pour l'égalité dx - presque partout des variables aléatoires définies sur $[0,1]$ à valeurs dans S telles que si X est un représentant d'une classe : $\exists\, s \in S$ tel que $E(d(X,s)) < +\infty$. Ce qui est alors vrai $\forall s \in S$, en utilisant l'inégalité triangulaire).
  On notera de façon abrégée L^1.
  L'espace L^1 est muni de la distance $E_\rho(d(X,Y)) = d_1(\tilde{X}, \tilde{Y})$ où $\tilde{X}$ est la classe d'équivalence de la variable aléatoire X.
  L'espace L^1 est alors un espace métrique complet.

- On notera $\mathcal{F} = \{ f / f : S \longrightarrow R$ lipschitzienne de rapport $1 \}$.

- On notera de la même façon l'espérance $E_\rho(d(X,Y))$ prise sur $S \times S$ par rapport à la loi conjointe ρ de (X,Y) et l'espérance par rapport à la mesure de Lebesgue sur $[0,1]$ de $d(X(\omega), Y(\omega)) : [0,1] \longrightarrow R$, par $E(d(X,Y))$.

34. <u>Lemme</u> : <u>Soit</u> X <u>une variable aléatoire de loi</u> P, Y <u>une variable aléatoire de loi</u> Q. <u>Alors</u> :

$$\| P-Q \|_L^* \leq E(d(X,Y))$$

<u>Preuve</u> : Soit $f \in \mathcal{F}$. Alors :

$$\int_S f\, d(P-Q) = \int_0^1 f(X(\omega))\, d\omega - \int_0^1 f(Y(\omega))\, d\omega$$

$$= \int_0^1 [\, f(X(\omega)) - f(Y(\omega)) \,]\, d\omega \leq \int_0^1 d(X(\omega), Y(\omega))\, d\omega$$

$$= E\,(d(X,Y)) \;.$$

Comme ce résultat est vrai pour toute $f \in \mathcal{F}$, on en déduit le lemme :

35. <u>Lemme</u> : <u>Soient</u> P <u>et</u> Q <u>deux probabilités sur</u> $(S, \mathcal{B})$ <u>à support fini. On peut construire deux variables aléatoires</u> X , <u>de loi</u> P , Y , <u>de loi</u> Q <u>définies sur</u> $[0,1]$ <u>à valeurs dans</u> S , <u>telles que</u> :

$$\|P - Q\|_L^* = E\,(d(X,Y)) \;.$$

<u>Preuve</u> : nous supposerons que $P = \sum_{i=1}^{\ell} p_i\, \delta_{x_i}$ où δ_{x_i} est la masse de Dirac en x_i , et $Q = \sum_{j=1}^{m} q_j\, \delta_{y_j}$, $p_i \geq 0$, $q_j \geq 0$.

On a : $\sum_{i=1}^{\ell} p_i = \sum_{j=1}^{m} q_j = 1$.

Soit $f \in \mathcal{F}$. Posons $f(x_i) = f_{1i}$ pour $i = 1, \ldots, \ell$ $f(y_j) = f_{2j}$ pour $j = 1, \ldots, m$. Puis définissons $I = \{1, \ldots, \ell\}$ et $J = \{1, \ldots, m\}$.

On obtient tout de suite : $\int f\, d(P-Q) = \sum_{i=1}^{\ell} f_{1i}\, x_i - \sum_{j=1}^{m} f_{2j}\, y_j$.

Si $\exists$ au moins un $i \in I$ et un $j \in J$ tels que $x_i = y_j$, on utilise alors la décomposition de Jordan-Hahn :

$$P = (P \wedge Q) + (P-Q)^+$$

$$Q = (P \wedge Q) + (P-Q)^- \;.$$

Posons $P_1 = (P-Q)^+$ et $Q_1 = (P-Q)^-$.

On voit facilement que P_1 et Q_1 sont orthogonales et que $\|P-Q\|_L^* = \|P_1-Q_1\|_L^*$.

On peut donc supposer que $\forall i \in I$ et $\forall j \in J$ $x_i \neq y_j$.

Appelons $d_{ij} = d(x_i, y_j)$; $f = (f_{1i}, i \in I ; f_{2j}, j \in J) \in \mathbb{R}^{m+\ell}$ et $W(f) = \sum_{i \in I} f_{1i} p_i - \sum_{j \in J} f_{2j} q_j$, qui est une forme linéaire sur $\mathbb{R}^{m+\ell}$.

Alors on sait que (I.3.3) 18 :

36. Maximiser la forme linéaire W, sous les contraintes :

$$\begin{cases} \forall i \in I, \ \forall j \in J, & |f_{1i} - f_{2j}| \leq d_{ij} \\ \forall i_1 \in I, \ i_2 \in I, & |f_{1i_1} - f_{1i_2}| \leq d_{i_1 i_2} \\ \forall j_1 \in J, \ j_2 \in J, & |f_{2j_1} - f_{2j_2}| \leq d_{j_1 j_2} \end{cases}$$

a pour dual

37. Minimiser sur ξ la valeur : $\sum_{\substack{i \in I \\ j \in J}} d_{ij} \xi_{ij}$ sous les contraintes :

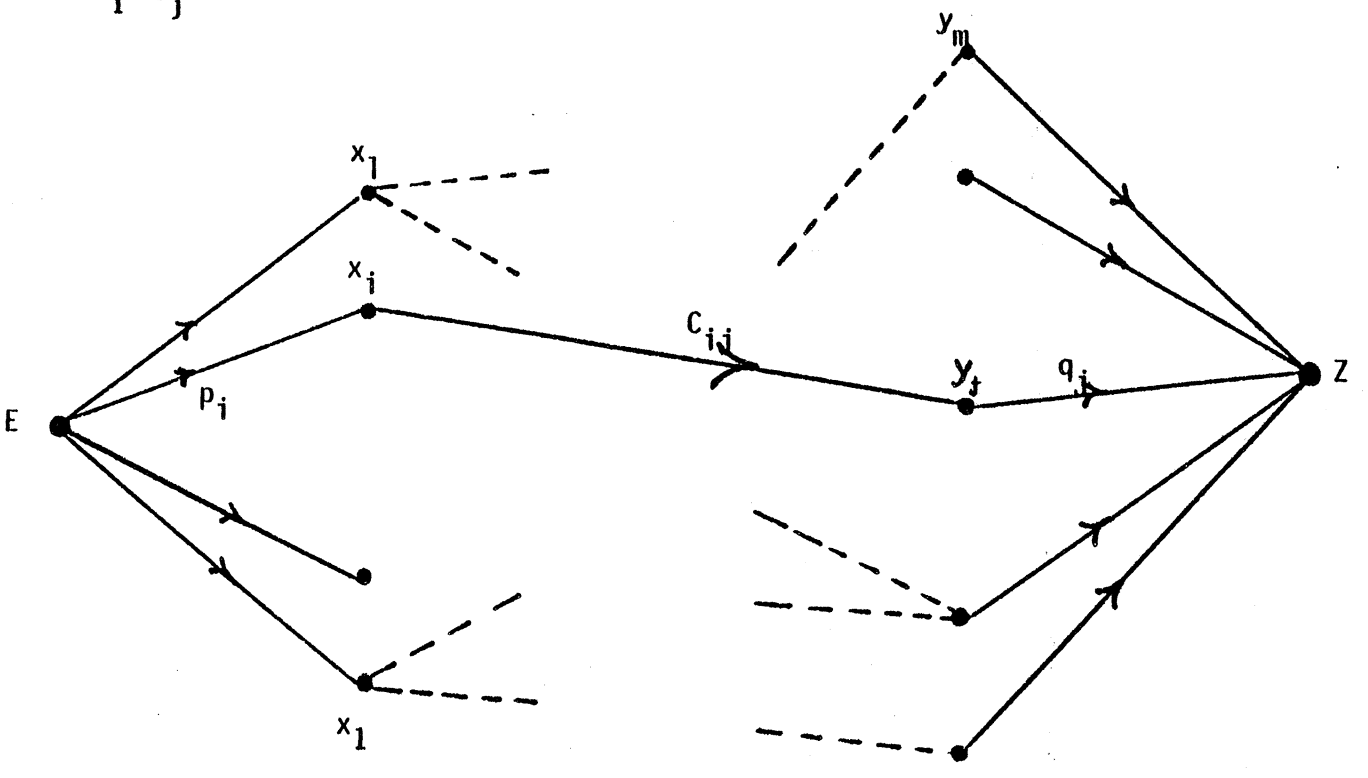
$$\begin{cases} \xi_{ij} \geq 0 \\ \sum_{i \in I} \xi_{ij} = q_j, \quad \sum_{j \in J} \xi_{ij} = p_i \end{cases}$$

qui est un problème de transport connu sous le nom de problème de Hitchcock.

Considérons le réseau de transport auxilaire suivant. Appelons $\{x_1, \ldots, x_\ell\}$ l'ensemble des sommets de production et $\{y_1, \ldots, y_m\}$ l'ensemble des sommets de consommation.

L'entrée E est reliée à x_i par un arc de capacité p_i. Le sommet x_i est relié à y_j par un arc de transit de capacité C_{ij} et y_j est relié à la sortie Z par un arc de capacité q_j.

Les arcs (E, x_i) sont appelés arcs initiaux.
Les arcs (y_j, Z) sont appelés arcs terminaux.
On cherche un flot ξ qui sature les arcs initiaux et terminaux et qui minimise la quantité $\sum_{i,j} d_{ij}\, \xi_{ij}$, ξ_{ij} étant la valeur du flot sur l'arc (x_i, y_j).



Rappelons le résultat suivant [2].

38. <u>Critère de Gale</u> : Une condition nécessaire et suffisante pour que le problème ait une solution est :

$$p_i \geq 0 \quad , \quad q_j \geq 0$$

$$\sum_{i=1}^{l} p_i = \sum_{j=1}^{m} q_j \quad \text{et}$$

$$\sum_{i=1}^{l} \max\Big(p_i \,,\, \sum_{j=j_1,\dots,j_k \subset J} C_{ij}\Big) \geq \sum_{j=j_1,\dots,j_k \subset J} q_j$$

Comme on peut supposer $C_{ij} = +\infty$ ou plus précisément $C_{ij} = q_j \quad \forall i \in I$, on voit que le problème dans notre cas est toujours possible. On a d'ailleurs utilisé au paragraphe I.3.3 un algorithme de calcul des ξ_{ij}.

Remarque : Au cas où l'on a recours à P_1 et Q_1 pour le calcul du flot ξ optimal, on voit que l'on peut revenir à P et Q en faisant circuler d'un sommet x_i vers lui-même la valeur $\min(P(x_i), Q(x_i))$.

La distance $d(x_i, x_i) = 0$, la valeur du coût minimal s'en trouve inchangée. De cette façon on aura toujours :

$$\sum_{\substack{i \in I \\ j \in J}} \xi_{ij} = 1 .$$

Plus précisément, on peut montrer que le problème de Hitchcock défini par

$$\{ \text{sommets de production} \} \cap \{ \text{sommets de consommation} \} = \phi$$

qui nous intéresse peut se ramener à un problème de transfèrement (intersection non vide) ; et réciproquement.

De plus les coûts étant donnés par une distance la solution du problème de transfèrement avec pour $y_j = x_i$: $\xi_{ij} = \min(P(x_i), Q(x_i))$ est la même que celle du problème de Hitchcock.

Posons alors

$$X(\omega) = \begin{cases} x_1 & \text{sur } [0, p_1[\\ x_2 & \text{sur } [p_1, p_1+p_2[\\ \vdots & \\ x_\ell & \text{sur } [\sum_{i=1}^{\ell-1} p_i, 1] \end{cases}$$

$$Y(\omega) = y_1 \quad \text{sur} \quad [0, \xi_{11}[\cup [p_1, p_1+\xi_{21}[\cup \ldots \cup [\sum_{i=1}^{\ell-1} p_i \,;\, \sum_{i=1}^{\ell-1} p_i + \xi_{\ell 1}[$$

$$Y(\omega) = y_2 \quad \text{sur} \quad [\xi_{11}, \xi_{11}+\xi_{12}[\cup [p_1+\xi_{21}, p_1+\xi_{21}+\xi_{22}[\cup \ldots \cup$$

$$[\sum_{i=1}^{\ell-1} p_i + \xi_{\ell 1}) \; \sum_{i=1}^{\ell-1} p_i + \xi_{\ell 1} + \xi_{\ell 2}[$$

$$\vdots$$

$$Y(\omega) = y_m \quad \text{sur} \quad [\sum_{j=1}^{m-1} \xi_{1j}, p_1[\cup [p_1 + \sum_{j=1}^{m-1} \xi_{2j}, p_1+p_2[\cup \ldots \cup$$

$$[1-p_\ell + \sum_{j=1}^{m-1} \xi_{\ell j}, 1] .$$

Alors $E(d(X,Y)) = \int_0^1 d(X(\omega), Y(\omega))\, d\omega = \sum_{i,j} d_{ij}\, \xi_{ij} = \|P-Q\|_L^* .$

On peut démontrer le lemme suivant : Bleuez [3]

39. <u>Lemme</u>. <u>Soit</u> (B_i) <u>pour</u> $i = 1, \ldots, \ell$ <u>une partition mesurable de</u> $[0,1]$. <u>Posons</u> $p_i = \int_{B_i} dx$ et $p_o = 0$.

<u>Soit alors</u> $D_i = [\sum_{j=0}^{i-1} p_j , \sum_{j=0}^{i} p_j[$ <u>pour</u> $i = 1, \ldots, \ell$.

<u>Alors il existe une variable aléatoire</u> T <u>de</u> $[0,1]$ <u>dans</u> $[0,1]$ <u>de loi</u> (dx) <u>telle que pour tout</u> $i \in I$, $T(B_i) \subseteq D_i$.

40. <u>Théorème</u> : <u>Soient</u> P <u>et</u> Q <u>deux lois de probabilité à support fini sur</u>$(S, \mathcal{B}$

<u>Alors</u> $\forall X$ <u>de loi</u> P , $\exists Y$ <u>de loi</u> Q <u>telle que</u>

$$\|P-Q\|_L^* = E(d(X,Y)) .$$

<u>Preuve</u> : Soit X une variable aléatoire définie sur $[0,1]$, $\mathcal{B}[0,1]$ de loi $P = \sum_{i=1}^{\ell} p_i \delta_{x_i}$. Posons $B_i = X^{-1}\{x_i\}$.

Soit alors $X_1 = \sum_{i=1}^{\ell} x_i \mathbb{1}_{D_i}$. On a $X = X_1 \circ T$.

D'après le lemme 12, $\exists Y_1$ de loi Q telle que

$E(d(X_1, Y_1)) = \|P-Q\|$. Posons $Y = Y_1 \circ T$. On a :
$E(d(X_1 \circ T, Y_1 \circ T)) = E(d(X_1, Y_1))$ . .

41. <u>Lemme</u> : <u>Soit</u> P <u>une loi de probabilité à support fini</u>. <u>Soit</u> Q <u>une loi de probabilité appartenant à</u> $\mathcal{P}_1(S)$. <u>Alors on peut trouver une variable aléatoire</u> X <u>de loi</u> P <u>telle que</u> :

$$\|P-Q\|_L^* = \inf \{ E(d(X,Y)) \ / \ Y \ \underline{\text{de loi}} \ Q \} .$$

<u>Preuve</u> : dans ce qui suit, pour tout sous-ensemble K de S on notera $\bar{K}$ l'adhérence de K

Compte tenu des hypothèses sur Q et sur S, étant donné $n \in \mathbb{N}$ et $y_0 \in S$ on peut trouver une partition mesurable finie : $K_{n_0}, \ldots, K_{n_1}$ telle que :

$$\begin{cases} \displaystyle\int_{k_{n_o}} (1 \vee d(y_o, x))\, dQ(x) < \frac{1}{2^3 n} \\ \\ \text{diam } K_{n_h} < 1/2^2 n \qquad \text{pour } h = 1,\dots,l \ . \end{cases}$$

Pour tout $h = 1,\dots,l$ on choisit un point $y_h \in K_{n_h}$ et l'on a :

$\forall\, x \in K_{n_h} \qquad d(x, y_h) < \dfrac{1}{2^2 n}$

Posons $\qquad Q_n = \sum_{h=0}^{l} \delta y_h \ Q(K_{n_h})$

On a d'après le lemme 11 page 69

$$||Q - Q_n||_L^* < 1/2n$$

Fixons provisoirement n.

$\forall\, h = 1,\dots,l \qquad \overline{K}_{n_h}$ est un sous ensemble compact de S.

On peut donc construire une partition fihie mesurable de K_{n_h} :

$K_{n_h n_1},\dots, K_{n_h n_{l_h}}$ de diamètre $< \dfrac{1}{2^3 n}$

Pour tout $i = n_1,\dots, n_{l_h}$

on choisit un point $y_{h\ i}$. Alors pout tout $x \in K_{n_{h\ i}}$ on a :

$d(x, y_{h\ i}) < \dfrac{1}{2^3 n}$

D'autre part l'ensemble $\overline{K}_{n_o}$ est un espace polonais.

Etant donné $y_{oo} \in K_{n_o}$, on peut trouver une partition mesurable finie de K_{n_o} notée $K_{n_o n_o}, \ldots, K_{n_o n_{l_o}}$ telle que :

$$- \int_{K_{n_o n_o}} [1 \vee d(y_{oo},x)] \, dQ(x)$$

$$\leq \int_{\overline{K}_{n_o n_o}} [1 \vee d(y_{oo},x)] \, dQ(x) < \frac{1}{2^4 n(d(y_o,y_1) \vee 1)}$$

$$- \operatorname{diam} K_{n_o n_h} < \frac{1}{2^3 n}$$

Choisissons pour chacun d'eux un point y_{oi} pour $i = 1,\ldots,l_o$ tel que $d(y_o,y_{oi}) < \inf_{K_{n_o n_i}} d(y_o,x) + \varepsilon$ où ε pourra être une quantité arbitrairement petite mais fixée.

Posons $Q_n^1 = \sum_{h=0}^{l} \sum_{i=0}^{l_h} Q(K_{n_h n_i}) \, \delta_{y_{hi}}$.

En utilisant le lemme 11 page 69 on obtient :

$$||Q - Q_n^1||_L^* < \frac{1}{2^2 n}$$

Appelons X la variable aléatoire de loi P construite au lemme 35 et qui est maintenant fixée. Alors toujours d'après le lemme 35, pour tout n on peut trouver une variable aléatoire Y_n de loi Q_n définie sur $[0,1]$ telle que $\|P - Q_n\| = E(d(X,Y_n))$.

A partir de la variable aléatoire Y_n nous allons construire la variable aléatoire Y_n^1 de la façon suivante : répartissons le flot optimal ξ_{ih} qui va de x_i vers y_h en $\xi_{ih1},\dots,\xi_{ihl_h}$ tel que

$$\sum_{k=0}^{l_h} \xi_{i\,h\,k} = \xi_{i\,h} \quad \text{et tel que}$$

$$\sum_{i=1}^{l} \xi_{i\,h\,k} = Q(K_{n_h n_k})$$

Supposons que la construction du lemme 37 soit celle de Y_n(aux indices près).

on a donc $Y_n(\omega) = y_h$ sur

$$\left[\sum_{j=0}^{h-1} \xi_{1j}\,,\ \sum_{j=0}^{h} \xi_{1j}\right[$$

$$\cup \left[p_1 + \sum_{j=0}^{h-1} \xi_{2j}\,,\ p_1 + \sum_{j=0}^{h} \xi_{2j}\right[$$

$$\cup \dots$$

$$\cup \left[\sum_{i=1}^{\ell-1} p_i + \sum_{j=0}^{h-1} \xi_{\ell j}\ ;\ \sum_{i=1}^{\ell-1} p_i + \sum_{j=0}^{h} \xi_{\ell j}\right[$$

posons alors :

$$Y_n^1(\omega) = y_{h1} \quad \text{sur} \quad \left[\sum_{j=0}^{h-1} \xi_{1j}\,,\ \sum_{j=0}^{h-1} \xi_{1j} + \xi_{1hj}\right[$$

$$\cup \left[p_1 + \sum_{j=0}^{h-1} \xi_{2j}\,,\ p_1 + \sum_{j=0}^{h-1} \xi_{2j} + \xi_{2h1}\right[$$

$$\cup \dots \cup \left[\sum_{i=1}^{\ell-1} p_i + \sum_{j=0}^{h-1} \xi_{\ell j}\,,\ \sum_{i=1}^{h-1} p_i + \sum_{j=0}^{h-1} \xi_{\ell j} + \xi_{\ell h1}\right[$$

.
.
.
.
.

$$Y_n^1(\omega) = y_{h\ell_h} \quad \text{sur} \quad \Big[\sum_{j=0}^{h-1} \xi_{1j} + \sum_{k=0}^{\ell_h-1} \xi_{1hk} \,,\, \sum_{j=0}^{h} \xi_{1j} \Big[$$

$$\cup \quad \Big[p_1 + \sum_{j=0}^{h-1} \xi_{2j} + \sum_{k=0}^{\ell_h-1} \xi_{2hk} \,,\, p_1 + \sum_{j=0}^{\ell} \xi_{2j} \Big[$$

$$\cup \ldots \cup \quad \Big[\sum_{i=1}^{\ell-1} p_i + \sum_{j=0}^{h-1} \xi_{\ell j} + \sum_{k=0}^{\ell_h-1} \xi_{\ell hk} \,,\, \sum_{i=1}^{\ell-1} p_i + \sum_{j=0}^{h} \xi_{\ell j} \Big[\; .$$

Alors $\displaystyle\int_{S-Kn_o} d(Y_n,Y_n^1)\, dQ(x) < \frac{1}{2^2 n} \left(1 - \frac{1}{2^3 n}\right)$

et $\displaystyle\int_{Kn_o} d(Y_n,Y_n^1)\, dQ(x) < \frac{1}{2^4 n} + \left(\varepsilon + \frac{1}{2^3 n}\right)\left(1 - \frac{1}{2^4 n}\right)$

On peut choisir ε de telle sorte que :

$$\left(\varepsilon + \frac{1}{2^3 n}\right)\left(1 - \frac{1}{2^4 n}\right) \leqslant \frac{1}{2^3 n}$$

donc $E\, d(Y,Y_n^1) < \dfrac{1}{2n}$.

On construit ainsi par récurrence une suite (Y_n^p) telle que si l'on note Q_n^p la loi de (Y_n^p) on a :

$$||Q - Q_n^p||_L^* < \frac{1}{2^{p+1} n}$$

et $\quad E\, d(Y_n^{p-1}, Y_n^p) < \dfrac{1}{2^p n}$

Donc $E\, d(Y,Y_n^p) \leq \displaystyle\sum_{l=1}^{p-1} E\, d(Y_n^l, Y_n^{l+1}) \leq \frac{1}{n}\left(\frac{1}{2} + \frac{1}{2^2} + \ldots + \frac{1}{2^p} + \ldots\right) < \frac{1}{n}$.

Donc la suite (Y_n^p) est dans L_1 et de Cauchy dans L_1. Donc (Y_n^p) converge en loi. Comme Y_n^p a pour loi Q_n^p et $\lim\limits_{p\to+\infty} Q_n^p = Q$ pour la convergence faible des probabilités, on en déduit que la loi de $\lim Y_n^p$ est Q. On note Y_n^Q cette variable aléatoire.

Alors $\forall n$: $E(d(X, Y_n^Q)) \leq E(d(X, Y_n)) + E(d(Y_n, Y_n^Q))$

$$\leq \|P - Q_n\|_L^* + \frac{1}{2n} \leq \|P - Q\|_L^* + \frac{2}{n} .$$

On en déduit à l'aide du lemme 36 que

$$\|P - Q\|_L^* = \inf \{ E(d(X, Y_n^Q)) \} .$$

42. <u>Corollaire</u> : <u>Soit</u> X <u>une variable aléatoire de</u> $[0,1]$ <u>dans</u> S <u>de loi</u> P <u>à support fini</u>. <u>Soit</u> Q <u>une loi de probabilité sur</u> $(S,d) \in \mathcal{P}_1(S)$. <u>Alors</u>

$$\|P - Q\|_L^* = \inf \{ E(d(X,Y)) \ / \ Y : [0,1] \to S \text{ de loi } Q \} .$$

<u>Preuve</u> : Soit X une variable aléatoire définie sur $[0,1]$, $\mathcal{B}[0,1]$ de loi $P = \sum_{i=1}^{\ell} p_i \, \delta_{x_i}$.

En utilisant le lemme (39) on a $X = X_1 \circ T$ où $X_1 = [\sum_{j=0}^{i-1} p_j , \sum_{j=0}^{i} p_j [$.

Alors d'après le lemme 41 pour tout ε on peut trouver une variable aléatoire Y_ε de loi Q telle que $E(d(X_1, Y_\varepsilon) \leq \|P - Q\|_L^* + \varepsilon$.

On en déduit que $E(d(X_1 \circ T , Y_\varepsilon \circ T)) = E(d(X_1, Y_\varepsilon)) \leq \|P - Q\|_L^* + \varepsilon$

$Y_\varepsilon \circ T$ a pour loi Q . On en conclut le corollaire.

43. <u>Corollaire</u> : <u>Soit</u> X <u>une variable aléatoire de loi</u> P , <u>élément quelconque de</u> $\mathcal{P}_1(S)$. <u>Soit</u> Q <u>une loi de probabilité appartenant à</u> $\mathcal{P}_1(S)$. <u>Alors</u> :

$$\|P-Q\|_L^* = \inf\{(E(d(X,Y)) \ / \ Y \text{ définie sur } [0,1], \text{ de loi } Q\}.$$

<u>Preuve</u> : on construit une suite X_n de loi P_n à support fini telle que $E(d(X_n,X)) \underset{n \to +\infty}{\longrightarrow} 0$.

D'après le lemme 34, on a alors $\|P-P_n\|_L^* \underset{n \to +\infty}{\longrightarrow} 0$.

D'après le corollaire 42 $\forall \epsilon$ on peut trouver une variable aléatoire Y_ϵ telle que

$$E(d(X_n,Y_\epsilon)) \leq \|P_n - Q\|_L^* + \epsilon \leq \|P-Q\|_L^* + 2\epsilon$$

$$E(d(X,Y_\epsilon)) \leq E(d(X,X_n)) + E(d(X_n,Y_\epsilon)) \leq \|P-Q\|_L^* + 3\epsilon .$$

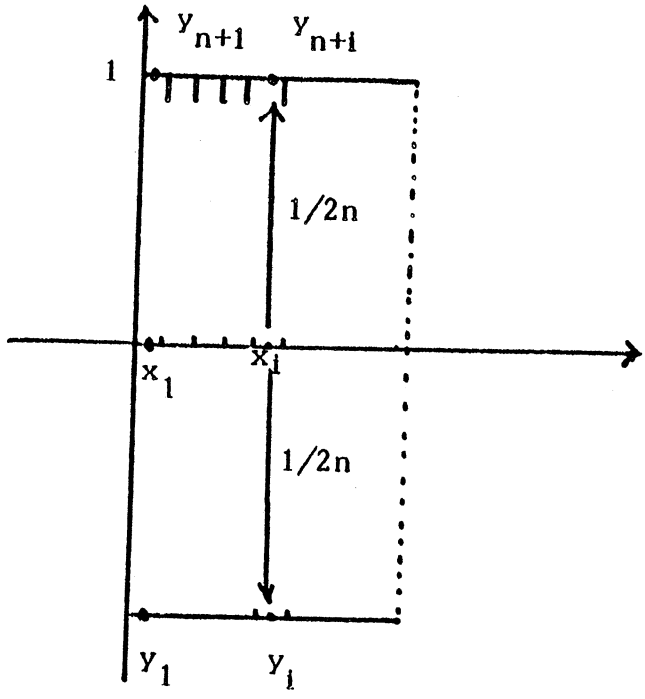
A l'aide du lemme 34 on conclut que

$$\|P-Q\|_L^* = \inf\{E\,d(X,Y) \ / \ Y \text{ de loi } Q\} .$$

43. <u>Exemples</u>.

Nous allons maintenant montrer sur un exemple qu'il existe des cas où l'inf n'est pas atteint et que donc le résultat ne peut pas être amélioré dans le cas général.

Considérons dans $\mathbb{R}^2$ muni de la norme euclidienne la loi P uniforme sur $[0,1] \times \{0\}$ et la loi Q uniforme sur $[0,1] \times \{1\} \cup \{0,1\} \times \{-1\}$.
Soit X l'application identique de $[0,1]$ dans $[0,1] \times \{0\}$.
L'application X est une variable aléatoire de loi P .
Effectuons alors une partition de $[0,1] \times \{0\}$ en n intervalles I_i $i = 1,\ldots,n$ de longueur $1/n$ et chargeons x_i milieu de I_i de la masse $1/n$.



Effectuons de même une partition de $[0,1] \times \{-1\}$ en n intervalles J_i $i = 1,\dots,n$ et de $[0,1] \times \{1\}$ en n intervalles J_i $i = n+1,\dots,2n$.
Appelons alors y_j $j = 1,\dots,2n$ le milieu de J_j qui est affecté de la masse $1/2n$.
On voit que $\min \sum_{i,j} x_{ij}\, d_{ij}$ où $d_{ij} = d(x_i, y_j)$ et x_{ij} le flot circulant de x_i vers y_j est obtenu en posant

$$x_{ii} = 1/2n = x_{i(n+i)}$$

$$x_{ij} = 0 \quad \text{pour} \quad j \neq i \quad \text{et} \quad j \neq n+i$$

et $\sum_{i,j} x_{ij}\, d_{ij} = 1$ (indépendant de n).

Par passage à la limite on voit que la variable aléatoire $Y(x)$ qui réaliserait le minimum devrait vérifier $Y(x) = (x,-1)$ et $Y(x) = (x,+1)$.
Donc Y n'existe pas.

Considérons la variable aléatoire X définie sur $[0,1]$ par

$$X(t) = \operatorname{Log} \frac{1}{1-t} .$$

Alors X est à valeurs dans $\mathbb{R}^+$ et P_X est définie par la densité $f(x) = e^{-x} \mathbb{1}_{\mathbb{R}^+}$.

Considérons alors la loi Q sur $\mathbb{R}^+$ définie par la densité $\frac{1}{2} e^{-x/2} \mathbb{1}_{\mathbb{R}^+}$.

On voit que $Y(t) = 2 \operatorname{Log} \frac{1}{1-t}$ est une variable aléatoire de loi Q définie sur $[0,1]$ et l'on a :

$$\int_0^1 |X(t) - Y(t)| \, dt = \int_0^1 \operatorname{Log} \frac{1}{1-t} \, dt$$

$$= u \operatorname{Log} u - u \Big|_1^0 = 1 = \|P-Q\|_L^* .$$

L'inf est donc atteint.

Plus généralement, en utilisant le théorème de Kantorovitch et Rubinstein et la formule de Wallender on voit que si $P \in \mathcal{P}_1(\mathbb{R})$ et si X est la variable aléatoire définie sur $[0,1]$ comme la pseudo-inverse de la fonction de répartition F de P , alors pour toute loi $Q \in \mathcal{P}(\mathbb{R})$ la variable aléatoire Y définie sur $[0,1]$ comme la pseudo-inverse de la fonction de répartition G de Q réalise l'infimum et l'on a :

$$\int_0^1 |X(t) - Y(t)| \, dt = \int_0^1 |F^{-1}(y) - G^{-1}(y)| \, dy$$

$$= \|P-Q\|_L^* .$$

BIBLIOGRAPHIE du paragraphe I - 3.

[1] A. DE ACOSTA - Invariance principle in probability for triangular aways .... Annals of Probability(1982), vol. 10, n° 2, pp. 346 - 373

[2] C. BERGE, A. GHOUILA - HOURI - Programmes, jeux et réseaux de Transport - Dunod

[3] J. BLEUEZ - Nouvelle démonstration du théorème de Strassen - Pub. I.R.M.A. - Lille (1981), vol. 3 - II 1 - 21

[4] R.L. DOBRUSHIN - Prescribing a system of random variables by conditional distributions - T.P.A. - vol. XV N° 3 (1970) pp. 458 - 486

[5] R.M. DUDLEY - Convergence of Baire Measures. Studia mathematica, (1966), T. XXVII, pp. 251 - 268

[6] R.M. DUDLEY - Distances of probability measures and random variables. Annals of Math. Statistics, (1968), vol. 39 n° 5, pp. 1563 - 1572

[7] R.M. DUDLEY - Probabilities and metrics (1976) - Lecture notes séries, n° 45 - Aarhus Universitet

[8] R.M. DUDLEY - Translation of the Kantorovitch and Rubinstein theorem - Communication personnelle

[9] R.M. DUDLEY - J. NEVEU - On Kantorovitch and Rubinstein theorems (1980) - Communication personnelle

[10] R. FORTET - E. MOURIER - Convergence de la répartition empirique vers la répartition théorique - A.E.N.S. 70 (1953), pp. 266 - 285

[11] R. FORTET - Some aspects of analysis and probability - Wiley - (1958)

[12] P.J. HUBER - Robust Statistics - Wiley (1981)

[13] J.M. SIMONNARD - Programmation linéaire - Dunod 1962

[14] A. SKOROKHOD - Limit theorems for stochastic processes, T.P.A. N° 3, (1956), pp. 261 - 289

[15] A.M. VERSIK - Some remarks on the infinite dimensional problem of linear programming - Rus. M.S. N° 5 (1970) - pp. 117 - 124

[16] S. WALLENDER - Calculation of the Wasserstein distance between probabity distributions on the line - Theory of Probability and its Appl.

[17] L. WASSERSTEIN - Markov processes with countable state space describing large system of automata. Problemy Peredachi Informatsii 5 (1969) pp. 64 - 73

CHAPITRE II

INTRODUCTION : Ce second chapitre est consacré aux problèmes d'estimation. On s'intéresse d'abord aux propriétés asymptotiques. On montre, à l'aide d'un contre-exemple que la notion classique de normalité asymptotique de l'estimateur du maximum de vraisemblance ne se généralise pas de façon canonique à l'estimateur ensembliste du maximum de vraisemblance, dans le cas non régulier.

Puis on s'intéresse aux problèmes de fiabilité d'un estimateur relativement aux modèlesde base. Les hypothèses des modèles statistiques sont rarement vérifiées.Ce qui veut dire qu'il y a rarement équidistribution - souvent supposée gaussienne - ou indépendance. Les observations proviennent de lois mélangées. Parfois la variabilité intrinsèque des données entraine elle aussi la présence de valeurs aberrantes. Après un rappel de la méthodologie statistique en matière de test de détection de ces valeurs, on s'intéresse à la détection de valeurs aberrantes multidimensionnelles par visualisation. On termine par une brève introduction à la théorie de la robustesse.

Enfin on tente de procéder à un classement des estimateurs d'un paramètre de position. On retient un certain nombre d'échantillons connus sur lesquels on fait opérer divers estimateurs. A l'aide de plusieurs analyses en composantes principales, on montre qu'un facteur séparateur important est le poids qu'accorde chaque estimateur aux valeurs extrêmes.

II - 1 Un contre exemple pour l'étude de la normalité asymptotique de l'estimateur ensembliste du maximum de vraisemblance.

Afin de pouvoir étudier des cas où l'équation de vraisemblance n'a pas de solution unique, on introduit la notion d'estimateur ensembliste du maximum de vraisemblance (m.v.). Cette notion a été appliquée par divers auteurs pour étudier des structures statistiques log-concaves normales, dont nous rappelons ci-dessous la définition. Pour la notion d'intégrande convexe normal on pourra consulter [18].

1 - Définition : Nous dirons que la structure statistique $(\Omega, \mathcal{A}, \{P_\theta / \theta \in \Theta\})$ où Θ est une partie convexe d'un espace $\mathbb{R}^k$ est une structure log-concave normale si

a) la famille $\{P_\theta / \theta \in \Theta\}$ est dominée par une mesure σ-finie μ

b) il existe, pour tout $\theta \in \Theta$, une densité de probabilité $\phi(.,\theta)$ de P_θ par rapport à μ, telle que si l'on pose :

$$\forall\ \omega \in \Omega,\ \forall\ \theta \in \Theta \qquad f(\omega\ \theta) = -\log \phi(\omega,\theta)$$

$$\forall\ \omega \in \Omega,\ \forall\ \theta \in \mathbb{R}^k - \Theta \quad f(\omega,\theta) = +\infty$$

la fonction f soit un intégrande convexe normal sur $\Omega \times \mathbb{R}^K$.

2 - Exemple : on obtient un exemple d'une telle structure en posant $k = 1$, $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ et en prenant pour μ la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Soit $\Theta = \mathbb{R}$. On pose :

$$f(\omega,\theta) = \delta(\omega,[\theta,\theta+1]) = \begin{cases} 0 & \text{si } \theta \le \omega \le \theta+1 \\ +\infty & \text{sinon} \end{cases}$$

La loi P_θ associée est alors la loi uniforme sur $[\theta,\theta+1]$.

On remarque que le domaine de définition de la loi P_θ dépend de θ. On est donc dans un cas non régulier. D'autre part, pour ω_0 fixé on a $\forall\ \theta \in [\omega_0-1, \omega_0]$:

$$\frac{\partial}{\partial\theta} \log \phi(\omega_0,\theta) = 0$$

Soit Γ la multi-application définie sur $(\Omega,\mathcal{A})$ à valeurs dans l'ensemble des convexes fermés non vides de Θ par :

$$\Gamma(\omega) = \{\theta \in \Theta \ / \ \inf\{ f(\omega,\xi) \ / \xi \in \Theta\} = f(\omega,\theta)\}$$

Cette multiaplication, dont on montre la mesurabilité, est alors appelée estimateur ensembliste du m.v. Sous certaines hypothèses, qui son vérifiées dans le cas de l'exemple 2, on obtient la convergence presque sûre de l'estimateur ensembliste du maximum de vraisemblance.

Soit P_{θ_0} la loi uniforme sur $[\theta_0, \theta_0+1]$.

Soit $\omega_1,\ldots,\omega_n$ un échantillon empirique d'une telle loi. Alors

$$\sum_{i=1}^{n} f(\omega_i,\theta) = \begin{cases} 0 \ \text{ si } \ \sup\limits_{1\le i\le n} \omega_i - 1 \le \theta \le \inf\limits_{1\le i\le n} \omega_i \\ +\infty \text{ ailleurs.} \end{cases}$$

En conséquence, pour toute suite $\{\omega_n\}$ telle que $\forall\, n \in \mathbb{N}$ $\omega_n \in [\theta_0, \theta_0+1]$, et pour tout n, l'ensemble $\Gamma_n(\omega_1,\ldots,\omega_n)$ est non vide et l'on a :

$$\Gamma_n(\omega_1,\ldots,\omega_n) = [\sup_{1\le i\le n} \omega_i - 1, \ \inf_{1\le i\le n} \omega_i]$$

Il est bien connu que $\sup\limits_{1\le i\le n} \omega_i$ est un estimateur convergent $P_{\theta_0}^{\mathbb{N}^\infty}$ - presque sûrement vers θ_0+1 et que $\inf\limits_{1\le i\le n} \omega_i$ est un estimateur convergent $P_{\theta_0}^{\mathbb{N}^\infty}$ - presque sûrement vers θ_0.

Nous désirons maintenant connaître le comportement asymptotique d'un élément particulierde cet estimateur ensembliste du m.v. de façon à savoir s'il est possible d'obtenir dans ce cas la normalité asymptotique, par analogie avec le cas régulier. On ne perd pas de généralité en supposant $\theta_0 = 0$.

En utilisant les résultats obtenus par Gnedenko [19] et qu'on pourra aussi trouver dans David [20] on voit qu'en ce qui concerne les statistiques d'ordre de la loi P_0 on est dans le domaine d'attraction de la famille de lois Λ_2.

Pour $\omega_{(1)} = \inf \omega_i$, leur fonction de répartition est donnée par :

$$\Lambda_2(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp(-x^{\alpha}) & x > 0, \alpha > 0 \end{cases}$$

En effet soit F la fonction de répartition de la loi uniforme sur [0,1].

$$F(x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ x & \text{si} \quad 0 < x < 1 \\ 1 & \text{si} \quad x \geq 1 \end{cases}$$

et l'on a : $F(1) = 1$ et $\forall\, \varepsilon > 0 \quad F(1-\varepsilon) < 1$

$$\text{et} \quad \lim_{x \to 0^+} \frac{1 - F(1 - xk)}{1 - F(1-x)} = \lim_{x \to 0^+} \frac{kx}{x} = k.$$

Ici $\alpha = 1$.

La fonction de répartition de la loi $\inf\limits_{1 \leq i \leq n} \omega_i$ est égale à $1 - (1-x)^n$.

On en déduit que la loi de $\frac{n}{\sqrt{2}} \inf\limits_{1 \leq i \leq n} \omega_i$, à valeurs dans $[0, \frac{n}{\sqrt{2}}]$, admet pour fonction de répartition :

$$1 - \left(1 - \frac{\sqrt{2}\, x}{n}\right)^n .$$

En faisant tendre n vers $+\infty$ on en déduit que la loi asymptotique normalisée de $\inf\limits_{1 \leq i \leq n} \omega_i$ admet pour fonction de répartition

$$G(x) = 1 - e^{-\sqrt{2}\, x} \quad \mathbb{1}_{\mathbb{R}^+}(x) \quad \text{et pour densité :}$$

$$g(x) = \sqrt{2}\, e^{-\sqrt{2}\, x} \quad \mathbb{1}_{\mathbb{R}^+}(x)$$

Par des considérations évidentes de symétrie, on en déduit que la loi asymptotique de $\frac{n}{\sqrt{2}} \sup\limits_{1 \leq i \leq n} \omega_i - 1$ admet pour densité $k(x) = \mathbb{1}_{\mathbb{R}^-}(x)\, \sqrt{2}\, e^{\sqrt{2}\, x}$

Nous allons considérer maintenant la variable aléatoire :

$$W_n = \frac{\sup\limits_{1 \leq i \leq n} \omega_i - 1 + \inf\limits_{1 \leq i \leq n} \omega_i -}{2}$$

milieu du segment de l'estimateur ensembliste du m.v.

Calcul de la densité et des moments de W_n

Densité :

posons $\begin{cases} U_n = \inf_i \omega_i \\ V_n = \sup_i \omega_i \end{cases}$ et cherchons la loi du couple (U_n, V_n).

Calculons donc la fonction de répartition de ce couple. On a

$$P[U_n<u,\ V_n<v] + P[u\leq U_n,\ V_n<v] = P[V_n<v]$$

$$\text{or } P[u\leq U_n,\ V_n<u] = (v-u)^n \quad \text{et} \quad P[V_n<v] = v^n.$$

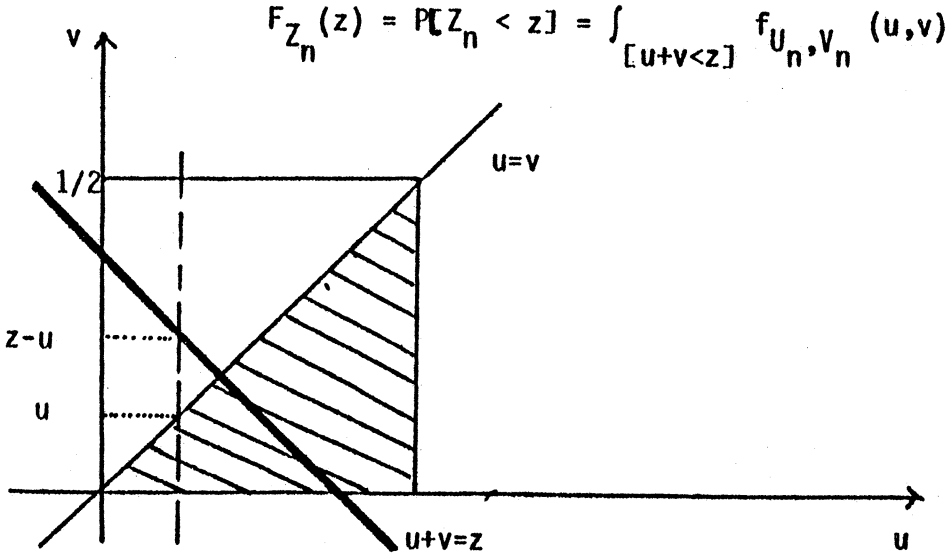
$$\text{D'où } F_{U_n,V_n}(u,v) = v^n-(v-u)^n \quad \text{si } 0<u<v<1.$$

La densité du couple (U_n,V_n) vaut :

$$f_{U_n,V_n}(u,v) = \frac{\partial^2}{\partial u\,\partial v} F_{U_n,V_n} = \begin{cases} n(n-1)(v-u)^{n-2} & \text{si } 0\leq u<v\leq 1 \\ 0 & \text{ailleurs} \end{cases}$$

Par intégration nous allons calculer la fonction de répartition de $Z_n = U_n+V_n$. On a

$$F_{Z_n}(z) = P[Z_n < z] = \int_{[u+v<z]} f_{U_n,V_n}(u,v)\,du\,dv$$



Distinguons deux cas : $0 \leq z \leq 1$ puis $1 \leq z \leq 2$.

a) $0 \leq z \leq 1$

(2) $$F_{Z_n}(z) = \int_0^{z/2} du \int_u^{z-u} n(n-1)(v-u)^{n-2}\,dv = \frac{z^n}{2}$$

b) $1 \leq z \leq 2$. On a

$$P[Z_n < z] = F_{Z_n}(z) = 1 - P[Z_n > z]$$

$P[Z_n > z]$ s'obtient à cause de la symétrie à l'aide de la formule (2) ci-dessus en remplaçant z par 2-z, soit :

$$F_{Z_n}(z) = 1 - \frac{1}{2}(2-z)^n$$

Posons alors $W_n = \frac{1}{2}(Z_n - 1)$.

$$F_{W_n}(w) = \begin{cases} \frac{(2w+1)^2}{2} & \text{si } -\frac{1}{2} \leq w \leq 0 \\ 1 - \frac{1}{2}(1-2w)^n & \text{si } 0 \leq w \leq \frac{1}{2} \end{cases}$$

Par dérivation

$$(3) \qquad f_{W_n}(w) = \begin{cases} n(2w+1)^{n-1} & \text{si } -\frac{1}{2} \leq w \leq 0 \\ n(1-2w)^{n-1} & \text{si } 0 \leq w \leq \frac{1}{2} \end{cases}$$

<u>moments</u> : en changeant w en -w dans l'expression de f_{W_n} pour $w \leq 0$ on obtient f_{W_n} pour $w \geq 0$. On en déduit que tous les moments d'ordre impair son nuls ; en particulier $E(W_1)$. Calculons les moments d'ordre pair.

$$E(W_n^{2p}) = 2\int_0^{1/2} n\, w^{2p}(1-2w)^{n-1} dw$$

posons $t = 1-2w$ $\quad dt = -2dw \quad$ et $\quad w = \frac{1}{2}(1-t)$

$$E(W_n^{2p}) = \frac{n}{2^{2p}}\int_0^1 (1-t)^{2p}\, t^{n-1}\, dt = \frac{n}{2^{2p}} B(n,2p+1) = \frac{n.(n-1)!(2p)!}{2^{2p}\,(n+2p)!}$$

$$\text{soit } E(W_n^{2p}) = \frac{2p!}{2^{2p}\,(n+1)(n+2)\ldots(n+2p)}$$

<u>Etude du comportement asymptotique</u> :

On a $\quad W_n \in [-1/2, +1/2]$

$$\sqrt{2}\, n\, W_n = \frac{n}{\sqrt{2}}\left(\sup_{1\leq i\leq n} \omega_i - 1\right) + \frac{n}{\sqrt{2}} \inf_{1\leq i\leq n} \omega_i$$

On a $\quad \sqrt{2}\, n\, W_n \in [-\frac{n}{\sqrt{2}}, +\frac{n}{\sqrt{2}}]$

et la loi de cette variable admet pour fonction de répartition :

$$\frac{1}{2}\left(\frac{\sqrt{2}w}{n} + 1\right)^n \quad \text{si } -\frac{n}{\sqrt{2}} \leq w \leq 0$$

$$1 - \frac{1}{2}\left(1 - \frac{\sqrt{2}w}{n}\right)^n \quad \text{si } 0 \leq w \leq \frac{n}{\sqrt{2}}$$

en faisant tendre n vers $+\infty$ on voit que la fonction de répartition limite est donnée par :

$$\begin{cases} \frac{1}{2}\, e^{\sqrt{2}\, w} & \text{si} \quad w \leq 0 \\ 1 - \frac{1}{2}\, e^{-\sqrt{2}\, w} & \text{si} \quad w \geq 0 \end{cases}$$

La densité correspondante est égale en définitive à

$$\frac{1}{\sqrt{2}}\, e^{-\sqrt{2}|\hat{w}|}$$

<u>Lemme</u> : <u>la densité de la loi asymptotique de la somme</u> :

$$\frac{n}{\sqrt{2}} \sup_{1\leq i\leq n} \omega_i - 1 + \frac{n}{\sqrt{2}} \inf_{1\leq i\leq n} \omega_i$$ <u>est égale au produit de convolution des densités des lois asymptotiques de</u> $\frac{n}{\sqrt{2}} \sup_{1\leq i\leq n} \omega_i - 1$ <u>et de</u> $\frac{n}{\sqrt{2}} \inf_{1\leq i\leq n} \omega_i$.

Calculons

$$k * g(w) = \int_{-\infty}^{+\infty} k(x)\, g(w-x)\, dx = 2 \int_0^{+\infty} e^{-\sqrt{2}\, x}\, e^{-\sqrt{2}(-(w-x))}\, 1_{\{(w-x)<0\}}\, dx$$

a) si $w < 0$ $\quad w-x < 0$

$$k * g = 2\, e^{\sqrt{2}\, w} \int_0^{+\infty} e^{-2\sqrt{2}\, x}\, dx = \frac{1}{\sqrt{2}}\, e^{\sqrt{2}\, w}$$

b) si $w \geq 0$

$$k * g = 2\, e^{\sqrt{2}\, w} \int_w^{+\infty} e^{-2\sqrt{2}\, x}\, dx = 2\, e^{\sqrt{2}\, w} \left(-\frac{1}{2\sqrt{2}}\, e^{-2\sqrt{2}\, x}\right)_w^{+\infty} = \frac{1}{\sqrt{2}}\, e^{-\sqrt{2}\, w}$$

d'où le résultat.

Plus généralement, prenons un point arbitraire de $[\sup_{1\leq i\leq n} \omega_i - 1, \inf_{1\leq i\leq n} \omega_i]$

Un tel point est caractérisé par

$$\alpha(\inf_{1\leq i\leq n} \omega_i) + (1-\alpha) \sup_{1\leq i\leq n} (\omega_i - 1)$$

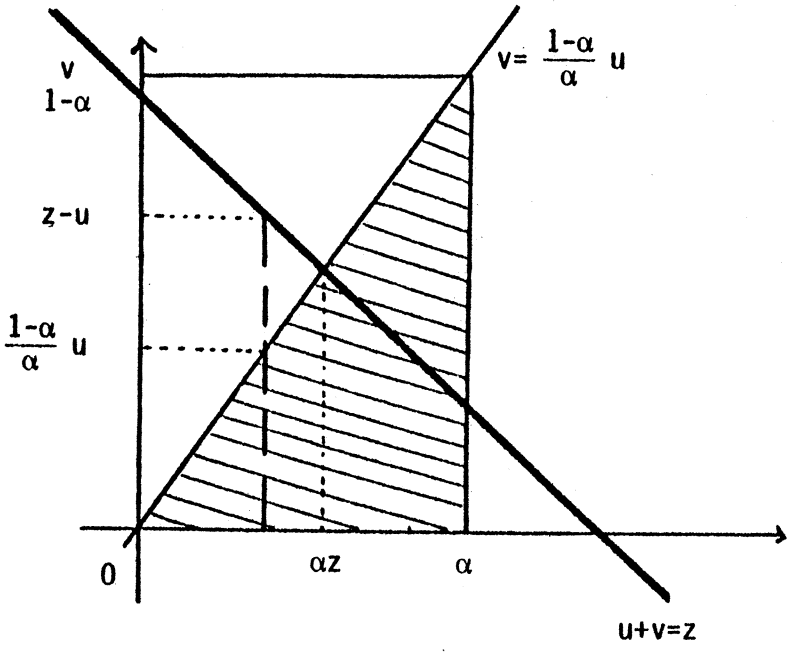
où $\alpha \in]0,1[$.

On se propose donc de faire un calcul analogue au calcul précédent en prenant cette fois

$$\alpha U_n + (1-\alpha) V_n \qquad 0 < \alpha \leq 1/2$$

On trouve que la densité du couple αU_n, $(1-\alpha)V_n$ est donnée par

$$f_{(U_n,\ (1-\alpha)V_n)}(u,v) = \begin{cases} \frac{n(n-1)}{\alpha(1-\alpha)} \left(\frac{v}{1-\alpha} - \frac{u}{\alpha}\right)^{n-2} & \text{si } 0 \leq \frac{u}{\alpha} \leq \frac{v}{1-\alpha} \leq 1 \\ 0 & \text{ailleurs} \end{cases}$$



On cherche alors la loi de $Z_n = \alpha U_n + (1-\alpha)V_n$.

Par intégration on voit que si $0 \leq z \leq 1-\alpha$

la fonction de répartition $G_{Z_n}(z) = P[Z_n < z]$

$$= \int_0^{\alpha z} du \int_{\frac{1-\alpha}{\alpha}u}^{z-u} \frac{n(n-1)}{\alpha(1-\alpha)} \left(\frac{v}{1-\alpha} - \frac{u}{\alpha}\right)^{n-2} dv$$

$$= (1-\alpha) \left(\frac{z}{1-\alpha}\right)^n \qquad \text{et que si} \quad 1-\alpha \leq z \leq 1$$

$$G_{Z_N}(z) = 1 - \alpha \left[\frac{1-z}{\alpha}\right]^n$$

On en conclut que la fonction de répartition $H_n(y)$ de la variable aléatoire $n(\alpha U_n+(1-\alpha)V_n-1+\alpha)$ à valeurs dans $[n(\alpha-1),n\alpha]$ est donnée par :

$$H_n(y) = \begin{cases} (1-\alpha)(1+\frac{y}{n(1-\alpha)})^n & \text{si } n(\alpha-1) \leq y \leq 0 \\ 1-\alpha(1-\frac{y}{n\alpha})^n & \text{si } 0 \leq y \leq n\alpha \end{cases}$$

en faisant tendre n vers $+\infty$ on voit que la fonction de répartition de la loi limite est donnée par

$$H(y) = \begin{cases} (1-\alpha)\, e^{\frac{y}{1-\alpha}} & \text{si} \quad y \leq 0 \\ 1-\alpha\, e^{-\frac{y}{\alpha}} & \text{si} \quad y \geq 0 \end{cases}$$

On n'obtient donc jamais la normalité asymptotique.

§ II - 2 DETECTION DES VALEURS ABERRANTES ET ROBUSTESSE

Un des problèmes qui se posent fréquemment dans l'analyse des données expérimentales est celui du traitement de quelques observations que leur taille - trop grande ou trop petite - rend suspecte. Ces observations douteuses, qui peuvent être le résultat de la variabilité des données ou plus souvent d'erreurs doivent être détectées : soit pour l'épuration des données, soit pour contrôler la fiabilité de certaines simulations ou de certaines méthodes de collecte des données.

La présence de valeurs aberrantes dans un échantillon a été modélisée par Grubbs et Dixon cf [13] ou [6] . Soient $X_1 \ldots X_n$ n variables aléatoires indépendantes :

Modèle $\mathcal{A}$: on suppose que n-k observations sont de loi $\mathcal{N}(\mu,\sigma^2)$ et $k < [n/2]$ sont de lois $\mathcal{N}(\mu+a_i, \sigma^2)$ $i = 1,\ldots,k$ où $a_i \in \mathbb{R}$ est inconnu.

Modèle $\mathcal{B}$: on suppose toujours que n-k observations sont de loi $\mathcal{N}(\mu\ \sigma^2)$ et k de loi $\mathcal{N}(\mu+b_i,\sigma^2)$ $i = 1,\ldots,k$ $b_i > 1$.

Le nombre k peut-être connu ou inconnu et représenter le nombre maximal de valeurs aberrantes. Diverses règles de rejet ont été proposées, basées sur des test ; et leur performances ont été étudiées relativement aux modèles $\mathcal{A}$ ou $\mathcal{B}$.

Dans le cas multidimensionnel, il est pratiquement impossible de se rendre compte, par une simple inspection des données, de la présence ou non de valeurs douteuses. Rohlf [14] écrit : (traduction faite par la rédaction)

"Malgré la complexité apparente du problème, on peut encore caractériser les valeurs aberrantes par le fait qu'elles se distinguent quelque peu du nuage de points principal. Peut-être ne s'isoleront-elles pas aux extrémités de la distribution comme dans le cas unidimensionnel, mais elles devront se singulariser quelque part. Des points qui ne sont pas à l'extérieur du nuage de points... sont, en puissance, des valeurs aberrantes. Il peut être utile de posséder des techniques pour déterminer la position d'un point relativement aux autres..."

Divers auteurs ont proposé par des méthodes de représentation graphique de données multidimensionnelles pouvant répondre aux préocupations de Rohlf (cf [4]). C'est l'une de ces méthodes que nous allons étudier.

II 2 - 1 Soit $x_1, \ldots, x_n$ un n-échantillon de dimension p : pour tout $i = 1, \ldots, n$ $x_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \end{pmatrix}$ Nous noterons x_i' le transposé de x_i.

Andrews [1] , [2] associe à chaque x_i le graphe entre $-\pi$ et $+\pi$ de la fonction 2π - périodique

(4) $$f_{x_i}(t) = x_i^1/\sqrt{2} + x_i^2 \sin t + x_i^3 \cos t + x_i^4 \sin 2t + \ldots$$

on voit qu'en posant $||f_x||_{L_2}^2 = \int_{-\pi}^{+\pi} f_x^2(t)\, dt$ on a

(5) $$||f_x||_{L_2}^2 = \pi ||x||^2$$ où $||.||$ désigne la norme euclidienne de $\mathbb{R}^p$.

Cette propriété, qui est l'une des plus intéressantes citées par Andrews, nous incite à utiliser le tracé des fonctions f_{x_i} pour la détection d'observations aberrantes.

6 - <u>Remarque</u> : il n'est pas pratique d'étudier par cette méthode de gros échantillons ; les graphes superposés devenant rapidement intraduisibles.

7 - <u>Remarque</u> : lorsque la détection des valeurs aberrantes nécessite la prise en compte d'une métrique autre que la métrique euclidienne, cette méthode appliquée à un échantillon brut est inopérante. C'est le cas, par exemple des échantillons gaussiens multidimensionnels de matrice des variances covariance inconnue où la métrique utilisée le plus couramment est celle de Mahalanobis. Voir [6].

Nous avons donc essayé d'adapter cette méthode pour construire un test visuel de détection de valeurs aberrantes.

8 - Soit $x_1, \ldots, x_n$ un échantillon gaussien de $\mathbb{R}^p$ de moyenne m et de matrice des variances - variances Λ inconnues. Λ et m sont respectivment estimées sans biais par

(9) $$S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$$

(10) $$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Nous savons - voir [6] - que dans le cas d'une observation aberrante provenant soit d'un glissement en moyenne, soit d'une dilatation de la matrice Λ, la règle de détection optimale est associée à la statistique

(11) $$T^2_{(n)} = \max_{i=1,\dots,n} (x_i-\bar{x})' \, S^{-1}(x_i-\bar{x})$$

De façon analogue, la règle de Thompson généralisée consiste à regarder les diverses valeurs T_i^2 pour $i = 1,\dots, n$ et à considérer comme détectée, toute observation x_i pour laquelle T_i^2 dépasse une certaine valeur critique. Ces tests sont facilement programmables.

Nous savons d'autre part que les règles de décision associées sont invariantes par translation et invariantes par le groupe linéaire, ceci par construction. La matrice S^{-1} est symétrique, presque sûrement inversible, définie > 0. Nous savons alors qu'il existe une matrice D telle que $D' \, S^{-1} \, D = \mathcal{I}_p$.

La règle de Thompson généralisée consiste à trouver les points situés à l'extérieur de l'ellipsoïde d'équation

(12) $$(x-\bar{x})' \, S^{-1} \, (x-\bar{x}) = c\phi R$$ où $c\phi R$ est la valeur critique. Posons

$x-\bar{x} = \sqrt{c\phi R}\,(DY)$. Il vient

(13) $$y' \, D' \, S^{-1} \, DY = 1 \text{ soit } y'y = 1$$

qui est l'équation de la sphère unité de $\mathbb{R}^p$. La règle ci-dessus consiste donc à chercher les points Y_i pour lesquels $Y_i' Y_i > 1$ avec $Y_i = \dfrac{D^{-1}(x_i-\bar{x})}{\sqrt{c\phi R}}$

On s'intéressera donc au tracés de f_{Y_i}.

<u>Mise en place du test</u>. Soit Y' tel que $Y' = \lambda Y$ avec $\lambda > 1$.

alors $$f_{Y'}(t) = f_{\lambda Y}(t) = \lambda y_1/\sqrt{2} + \lambda y_2 \sin t + \dots$$
$$= \lambda \, f_Y(t)$$

Ceci signifie que pour $f_Y(t) > 0$ (resp. $f_Y(t) < 0$) le graphe de $f_{Y'}$ sera dessus (resp. au dessous) de celui de f_Y.

Cette remarque nous conduit a adopter comme zone visuelle de confiance le sous-ensemble de $\mathbb{R}^2$ délimité par l'enveloppe supérieure et par l'enveloppe inférieure des graphes des fonctions f_Y, avec $Y \in S_1(p)$ où $S_1(p)$ est la sphère unité de $\mathbb{R}^p$. Dès qu'en un point $t \in [-\pi, +\pi]$ le graphe d'une fonction f_Y sort de cette zone de confiance, l'observation correspondante est détectée. Il est clair que seuls des points pour lesquels : $Y'Y > 1$, seront détectés. Mais cette méthode permet-elle de les détecter tous ?

Etude du cas p = 2

Tracé automatique de la zone de confiance :

La zone de confiance est déterminée par les enveloppes supérieure et inférieure des graphes du cercle unité C_1.

Détermination de l'enveloppe supérieure :

Soit $-\pi/2 \leq t \leq \pi/2$ on veut trouver $-\pi \leq \theta \leq +\pi$ tel que

(14) $$\frac{\cos\theta}{\sqrt{2}} + \sin\theta.\sin t \quad \text{soit maximal.}$$

on pose $a = \frac{1}{\sqrt{2}}$, $b = \sin t$. Dérivons (14). Il vient :

(15) $-a \sin\theta + b \cos\theta$. On obtiendra un max. ou un min. si (15) = 0

soit en posant $u = \operatorname{tg}\frac{\theta}{2}$ et $b \neq 0$

(16) $b(1-u^2) - 2au = 0$ dont les solutions sont

$$u_1 = \frac{-a+\sqrt{a^2+b^2}}{b} \qquad u_2 = \frac{-a-\sqrt{a^2+b^2}}{b}$$

On obtient deux solutions :

$$\theta_1 = 2 \operatorname{Arc\,tg} u_1 \qquad \theta_2 = 2 \operatorname{Arc\,tg} u_2$$

dont on vérifie que si l'une donne un max. l'autre donne le min. On conserve la valeur du max.

Détermination de l'enveloppe inférieure : par raison de symétrie cette enveloppe est la symétrique par rapport à Ox de l'enveloppe supérieure.

Initialisation : pour $t = -\frac{\pi}{2}$ $\sin t = -1$ $\theta_{max} = -.955$ radian

$(1)_{max} = 1.225$

pour $t = 0$ $\sin t = 0$ $\theta_{max} = 0$ $(1)_{max} = 0.707$

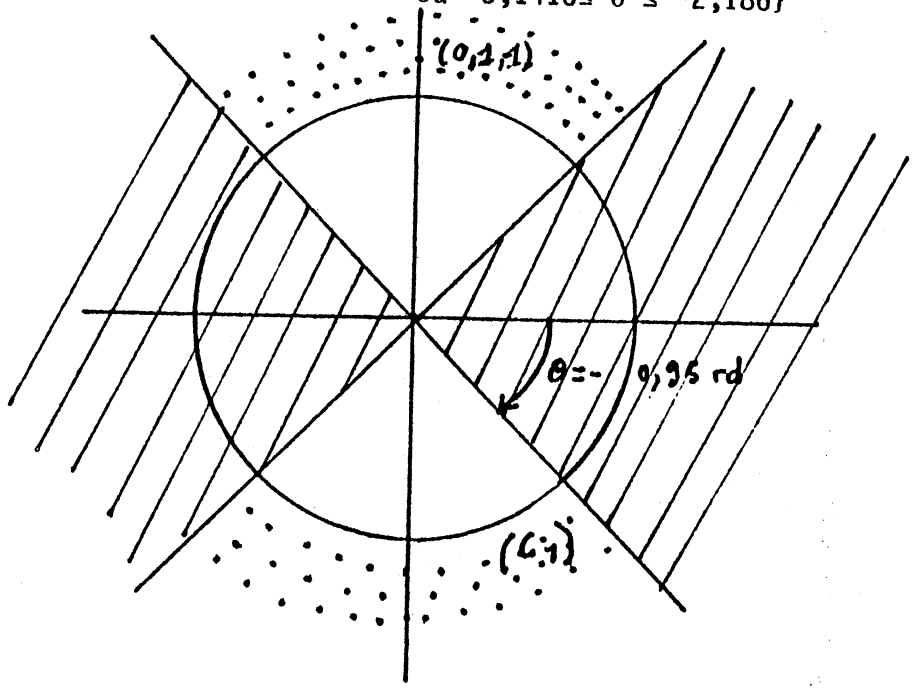
L'application $t \longrightarrow y(t)$ élément du cercle unité tel que $f_{y(t)}(t) \geq f_{y'}(t)$ $\forall\, y' \neq y$ élément du cercle unité est injective continue par morceaux. Malheureusement cette application n'est pas surjective. D'après les calculs sur ordinateur cette application établit une bijection entre

$$[-\pi/2, +\pi/2] \quad \text{et} \quad \{y \in C_1 / -0{,}95 \leq \theta \leq 0{,}95\}$$

les angles étant en radians.

De même l'application $t \longrightarrow y(t) \in C_1$ tel que $f_{y(t)}(t) \leq f_{y'}(t)$ $\forall\, y' \in C_1$ établit une bijection entre

$$[-\pi/2, +\pi/2] \quad \text{et} \quad \{y \in C_1 / +2{,}186 \leq \theta \leq +3{,}1416$$
$$\text{ou } -3{,}1416 \leq \theta \leq -2{,}186\}$$



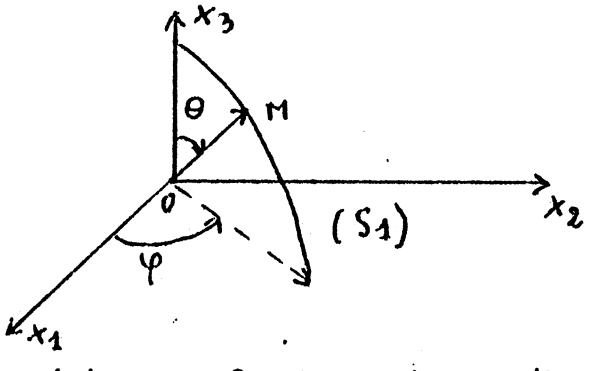
Ainsi des valeurs situées dans le secteur non-hachuré et relativement proches du cercle (C_1) tout en étant à l'extérieur ne seront pas forcément détectées : ce que l'on vérifie facilement pour (0; 1,1) par exemple. Mais $\pi/4 < 0{,}95$. On voit alors qu'en faisant une <u>permutation</u> des coordonnées et en traçant également $f_y(t) = y_2/\sqrt{2} + y_1 \sin t$, toute observation vérifiant $y'y > 1$ sera détectée au moins une fois ; et l'on obtient ainsi l'identique visuel de la règle numérique de Thompson généralisée.

Le graphe Visual II représente le tracé de la zone de confiance ainsi que le graphe des fonctions f associées aux points $x_1 = (5.38; -21.79)$ $x_2 = (12.46 ; -0.75)$ et $x_3 = (9.43 ; -15.60)$ auxquels correspondent les points $y_1 = (0.845 ; -1,421)$ $y_2 = (0.442 ; 1.087)$ et $y_3 = (0.788 ; -0.542)$.

Etude du cas p = 3

La zone de confiance est déterminée par les enveloppes supérieure et inférieure des graphès de la sphère S_1.

Détermination de l'enveloppe supérieure : Soit $-\pi \leq t \leq +\pi$



$x_1 = \sin\theta.\cos\varphi$

$x_2 = \sin\theta.\sin\varphi$

$x_3 = \cos\theta$

On veut trouver $0 \leq \theta \leq \pi$ et $-\pi \leq \varphi \leq +\pi$ tels que

(17) $$\frac{x_1}{\sqrt{2}} + x_2 \sin t + x_3 \cos t \quad \text{soit maximal}$$

x_1, x_2, x_3 étant les coordonnées du point $M \in (S_1)$

On pose $a = \frac{1}{\sqrt{2}}$, $b = \sin t$, $c = \cos t$ et $f(\varphi,\theta) = (17)$. On a donc

(18) $$f(\varphi,\theta) = a \sin\theta \cos\varphi + b \sin\theta \sin\varphi + c \cos\theta$$

$$f'_\varphi = \sin\theta(-a \sin\varphi + b \cos\varphi)$$

$$f'_\theta = \cos\theta(a \cos\varphi + b \sin\varphi) - c \sin\theta$$

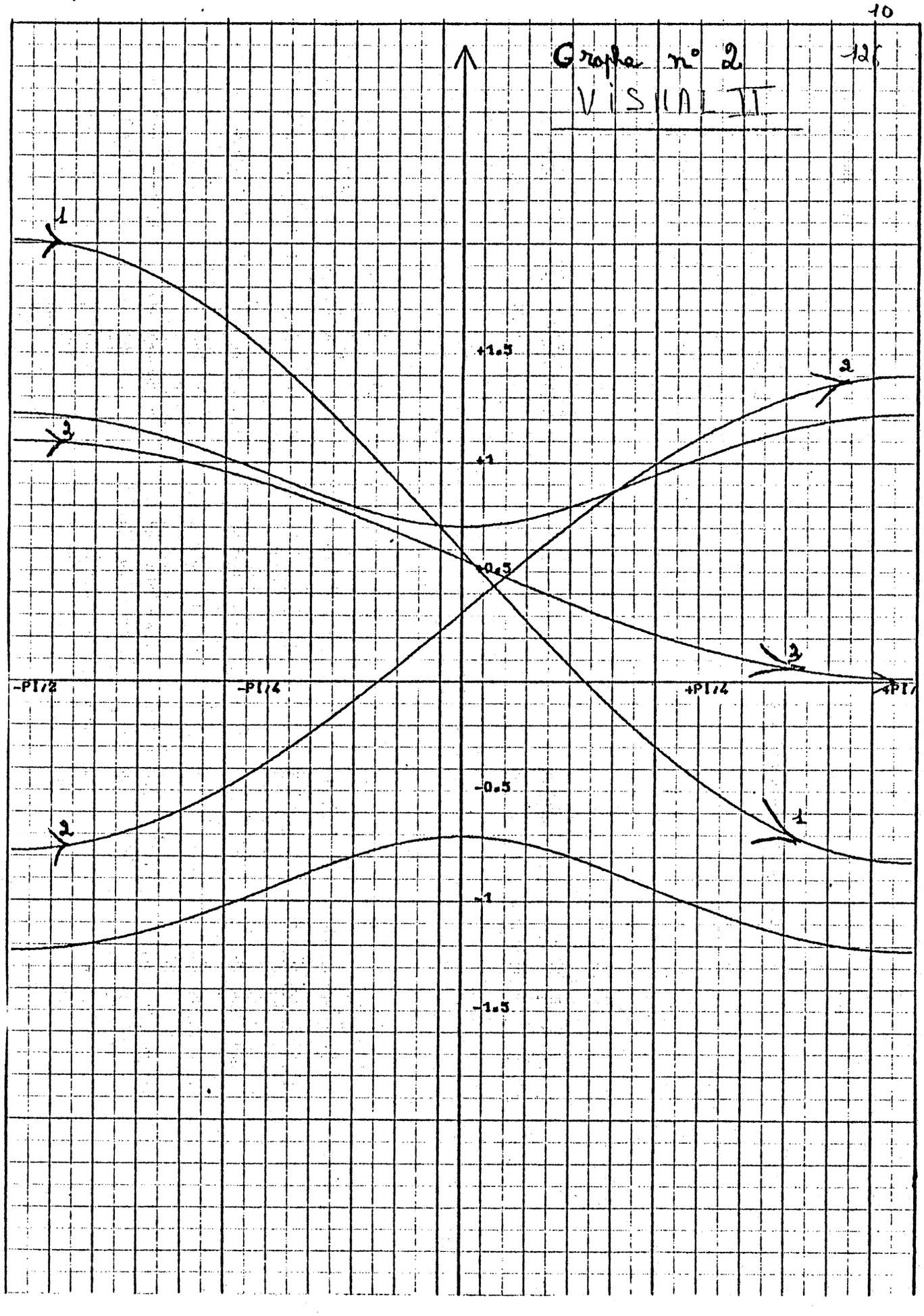
Les extremums de f sont obtenus en des points qui vérifient

$$f'_\varphi = f'_\theta = 0$$

1°) $$f'_\varphi = 0 \Leftrightarrow \begin{cases} \theta = k\pi \\ \text{ou} \quad -a \sin\varphi + b \cos\varphi = 0 \end{cases} \qquad (19)$$

posons $t = \operatorname{tg}\frac{\varphi}{2}$, $b \neq 0$. (19) s'écrit

$$-2at + b(1-t^2) = 0 \qquad (20)$$


Graphe n° 2
VISINLIT
+1.5
+1
+0.5
-0.5
-1
-1.5
-PI/2
-PI/4
+PI/4
1
2
3

dont les solutions sont : $t_1 = \frac{-a+\sqrt{a^2+b^2}}{b}$ $\quad t_2 = \frac{-a-\sqrt{a^2+b^2}}{b}$

Soit $\varphi_1 = 2 \text{ Arc tg } t_1$, $\varphi_2 = 2 \text{ Arc tg } t_2$

2°) $f'_\theta = 0$ posons $d = a \cos\varphi + b \sin\varphi \neq 0$

$$f'_\theta = 0 \Leftrightarrow d \cos\theta - c \sin\theta = 0 \qquad (21)$$

posons $u = \text{tg } \frac{\theta}{2}$ (21) s'écrit : $d(1-t^2) - 2ct = 0$

dont les solutions sont $u_1 = \frac{-c-\sqrt{c^2+d^2}}{d}$, $u_2 = \frac{-c+\sqrt{c^2+d^2}}{d}$

Soit $\theta_1 = 2 \text{ Arc tg } u_1$, $\theta_2 = 2 \text{ Arc tg } u_2$

La solution du 1° nous donne deux valeurs pour φ qui conduisent chacune à un couple de solutions (θ_1, θ_2). On vérifie que ces solutions sont le max. ou le min de (18)

<u>Détermination de l'enveloppe inférieure</u> : obtenue par symétrie par rapport à ox de l'enveloppe supérieure.

initialisation : $t = -\pi$

$$f(\varphi, \theta) = a \sin\theta \cos\varphi - \cos\theta$$

$$f'_\varphi = -a \sin\theta \sin\varphi = 0 \Leftrightarrow \begin{cases} \theta = k \\ \varphi = k\pi \end{cases}$$

$$f'_\theta = a \cos\theta \cos\varphi + \sin\theta = a \cos\theta + \sin\theta \quad \text{si } \varphi = 0$$

$$\theta_1 = 2 \text{ Arc tg } \frac{1+\sqrt{1+a^2}}{a} \quad , \quad \theta_2 = 2 \text{ Arc tg } \frac{1-\sqrt{1+a^2}}{a}$$

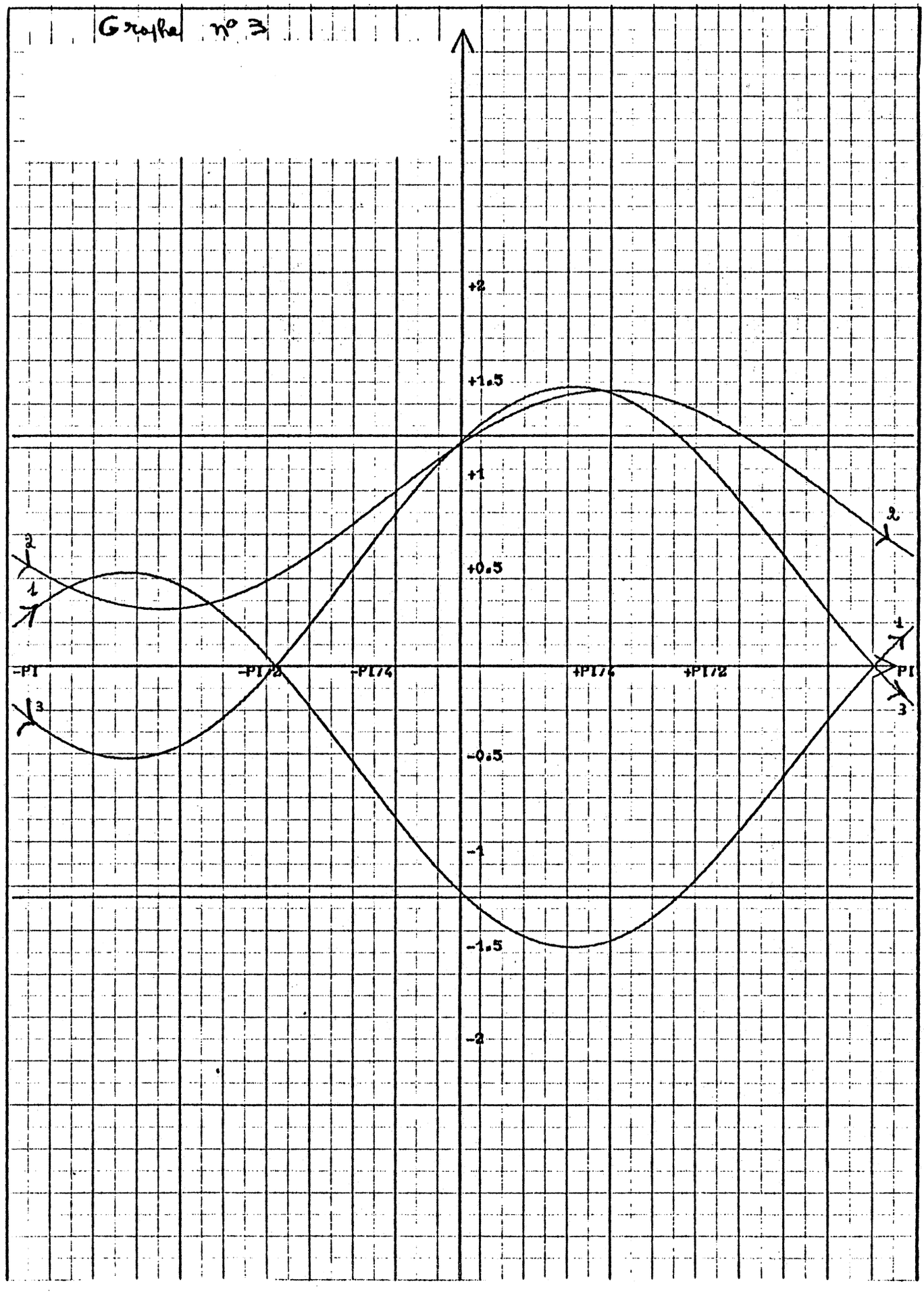
$\theta_1 = 2.526$ rd. et $f_{max} = 1.225$

pour $t = 0$, $\theta = 0$ $\varphi = 0{,}62$ rd $f_{max} = 1.225$

Cette fois il est hors de question d'obtenir par permutation des coordonnées le même résultat que pour p = 2. Le test associé va donc ignorer certaines valeurs situées en dehors de S_1.

Par exemple, le point $\begin{pmatrix} 0 \\ 0 \\ 1.1 \end{pmatrix}$ n'est jamais détecté.

Le graphe n° 3 représente la zone de confiance et les graphes associés aux points : $y_1 = (.7,.7,.7)$, $y_2 = (1.25,.5,.3)$ et $y_3 = (-.7, .7, .7)$


Graphe n° 3
+2
+1.5
+1
+0.5
-0.5
-1
-1.5
-2
-PI
-PI/2
-PI/4
+PI/4
+PI/2
PI
1
2
3

22 - Remarque : Posons

$$f_1(t) = \begin{pmatrix} 1/\sqrt{2} \\ \sin t \\ \cos t \\ \vdots \end{pmatrix}$$

Pour tout $t \in [-\pi, +\pi]$

(23) $$f_{x_i}(t) = x_i' . f_1(t) \quad \text{pour} \quad i = 1,\dots,n$$

Pour t fixé, on opère donc une projection du nuage de points $x_1,\dots x_n$ sur une direction de $\mathbb{R}^p$.

En faisant varier t , on fait varier la direction de projection et l'on pourrait espérer trouver ainsi la meilleure fonction discriminante (23), pour une observation x_i donnée. Malheureusement il n'y a pas de bijection entre l'ensemble des directions de $\mathbb{R}^p$ et l'ensemble de celles décrites par $f_1(t)$. Pour $p > 2$ l'ensemble des directions décrites par les vecteurs $f_1(t)$ ne vérifie aucune propriété de densité relativement à l'ensemble des directions de $\mathbb{R}^p$: toutes les composantes du vecteur f_1 sauf la première varient en même temps, en fonction de t. Il n'y a donc aucune raison de trouver dans le petit sous-ensemble des directions de $\mathbb{R}^p$ auquel on se limite, celle qui fournira une bonne fonction discriminante (23) pour une observation x_i donnée.

24 - Remarque : Signalons enfin que c'est la particularité évoquée en remarque 22 qui nous empêche de construire une zone de confiance visuelle qui permettrait de faire de la méthode d'Andrews un test rigoureux de détection de valeurs aberrantes dans certains échantillons.

II 2.2 - Introduction à la robustesse.

Lorsque le but de l'épuration des données est de procéder à une estimation, une méthodologie nouvelle s'est considérablement développée : on ne va plus procéder à des rejets de valeurs considérées comme aberrantes. On va rechercher de nouveaux estimateurs qui se comportent bien au voisinage des modèles théoriques

habituels contrairement aux estimateurs classiques dont on va illustrer plus bas l'instabilité. On calculera ces nouveaux estimateurs directement sur les données brutes. Cette méthode basée sur une modélisation moins figée de la notion d'aberration porte le nom de théorie de la Robustesse et vient d'être définitivement consacrée dans le livre de Huber [12].

Une des premières alternatives aux modèles de Grubbs et Dixon fut la suivante :

<u>Lois normales contaminées</u> : Soit P une loi normale de moyenne m et de variance σ^2 et Q une autre loi normale de moyenne m+a et de variance $b\sigma^2$ avec $b > 1$. On pose

(25) $$H = (1-\varepsilon)P + \varepsilon Q \quad \text{avec} \quad \varepsilon \in]0,1[.$$

La loi H est appelée noi normale contaminée. Proposée par Tukey J. [17] c'est la base d'un modèle maintenant très connu appelé modèle contaminé. On remarque que H n'est pas une loi normale et que H est une loi symétrique si et seulement si $\alpha = 0$. Le paramètre ε est le pourcentage de contamination ou plus simplement la contamination.

Dans un échantillon de taille n de la loi H, nous appelerons contaminants les observations qui suivent une loi Q. Leur nombre R est une variable aléatoire qui suit une loi binomiale de paramètres n et ε.

Ainsi les modèles $\mathcal{A}$ et $\mathcal{B}$ peuvent être considérés comme des modèles contaminés, avec pour $\mathcal{A}$: $b = 0$ et pour $\mathcal{B}$: $a = 0$, $\lambda = b$, où le nombre de contaminants est fixe. C'est pourquoi le modèle (25) est appelé par certains auteurs : "modèle de contamination aléatoire". Ce modèle paraît de prime abord plus satisfaisant pour prendre en compte la présence éventuelle de valeurs aberrantes puisque dans un échantillon, leur nombre est une variable aléatoire. Cependant il faut bien voir que l'on déplace l'inconnue de k vers ε , et la seule différence revient à considérer que cette fois c'est en moyenne qu'il y a $n\varepsilon$ observations aberrantes dans l'échantillon.

26 - Cas de la moyenne empirique.

Essayons d'évaluer l'instabilité. Supposons qu'une loi $P = \mathcal{N}(m,\sigma^2)$ est contaminée dans un pourcentage ε par une loi $Q = \mathcal{N}(m,b\sigma^2)$, $b > 1$. la moyenne empirique $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ est un estimateur sans biais de m pour un n-échantillon de P mais aussi de H définie en (25). Cependant, nous allons calculer la variance de $\bar{x}$ sous la loi H, notée $v_H(\bar{x})$. Nous savons que la variance d'une variable est égale à la somme de la variance liée moyenne et de la variance de la moyenne liée. Ici la variance de la moyenne liée est nulle donc :

$$v_H(\bar{x}) = E\,[E((\bar{x}-m)^2|R)] = \frac{1}{n^2}\,E\,[(n-R)\sigma^2 + R\,b\sigma^2]$$

or $E(R) = n\varepsilon$ d'où

(27) $$v_H(\bar{x}) = \frac{\sigma^2}{n}\,[1+(b-1)\varepsilon]\ .$$

Ce résultat soit être rapproché de $v_P(\bar{x}) = \frac{\sigma^2}{n}$: si nous estimons m à l'aide d'un échantillon du modèle contaminé, le facteur multiplicatif $1+(b-1)\varepsilon$ viendra modifier la varaince de notre estimateur. Pour illustrer ce phénomène, prenons $\varepsilon = 0{,}1$ et $b = 9$. Le facteur $1+(b-1)\varepsilon$ vaut alors 1,8 et nous enregistrons une perte d'efficacité absolue de 30%.

28 - Cas de la variance empirique.

Etudions maintenant les effets de la contamination sur les performances de la variance empirique pour estimer la variance.

Posons $$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

On sait que $E(s^2) = \frac{n-1}{n}\sigma^2$ et $v(s^2) = \frac{n-1}{n^3}\,[(n-1)\mu_4+(3-n)\sigma^4]$

où μ_4 est le moment centré d'ordre 4 de la loi de probabilité sous-jacente. Dans le cas d'une loi gaussienne de moyenne m de variance σ^2 on obtient $\mu_4 = 3\,\sigma^4$ soit :

(29) $$v(s^2) = \frac{2(n-1)\sigma^4}{n^2}.$$

Dans le cas du modèle contaminé on obtient facilement que

$$E(s^2/R) = \frac{n-1}{n^2}\sigma^2\,(n-R+Rb) \quad \text{et} \quad E(s^2) = \frac{n-1}{n}\sigma^2\,(1-\varepsilon+\varepsilon b).$$

Nous nous proposons de calculer $v_H(s^2)$

$v_H(s^2) = E(s^4) - E^2(s^2)$, où l'on peut écrire $E(s^4) = E(E(s^4/R))$. Comme

$$s^4 = \Big(\sum_{i=1}^{n}(x_i-m)^4\Big)\frac{(n-1)^2}{n^4} + \sum_{k=1}^{n}(x_k-m)^2\Big[\sum_{i>j}(x_i-m)(x_j-m)\Big]\Big(\frac{4}{n^4}-\frac{4}{n^3}\Big)$$
$$+\frac{(n-1)^2+4}{n^4}\sum_{i<j}(x_i-m)^2(x_j-m)^2 + \frac{8}{n^4}\sum_{\substack{i<j\\k<l\\(i,j)\neq(k,l)}}(x_i-m)(x_j-m)(x_k-m)(x_l-m).$$

On obtient

$$E(s^4/R) = \frac{3(n-1)^2}{n^4}\sigma^4[n+R(b^2-1)] + \sigma^4\frac{(n-1)^2+2}{n^4}.[(n-R)(n-R-1)+2(n-R)Rb+R(R-1)b^2]$$

et sachant que $E(R) = n\varepsilon$ et $E(R^2) = n^2\varepsilon^2 + n\varepsilon^2$, on trouve :

$$v_H(s^2) = \frac{2(n-1)^4}{n^2}\Big[1+\frac{1}{2}\varepsilon(b-1)(3b+1) - \frac{1}{2}\varepsilon^2(b-1)^2 - \frac{3\varepsilon(b-1)^2(1-\varepsilon)}{2n}\Big]$$

30 - <u>Remarque</u> : Les calculs précédents ne diffèrent que par le facteur $\frac{(n-1)^2}{n^2}$, qui s'ajoute, des calculs analogues concernant l'estimé sans biais de σ^2. La contamination a donc fait croître $v(s^2)$ du facteur entre crochets. Par exemple si $\varepsilon = 0{,}05$ et $b = 9$ ce facteur vaut $6{,}52 - \frac{4{,}56}{n}$. Si $n = 7$ on obtient 7,17 qui conduit inévitablement à une perte importante d'efficacité.

LES 3 MODELES.

31 - Le modèle contaminé.

Soit $\mathcal{M}$ une classe de lois assez vaste et $\varepsilon > 0$ voisin de 0 ; on définit la classe

$$\mathcal{F}(\varepsilon) = \{(1-\varepsilon)F_\theta+\varepsilon G \mid \theta\in\Theta \text{ et } G\in\mathcal{M}\}$$

$\mathcal{F}(\varepsilon)$ est la classe contaminée du modèle initial $\mathcal{F}(0) = \{F_\theta \mid \theta\in\Theta\}$.

Il s'agit s'une généralisation du modèle proposé par Tukey. Ce modèle a été utilisé, par exemple, par Huber P.J. [11] , pour démontrer des propriétés de types minimax.

32 - Le modèle à variation totale.

Si F et G sont deux probabilités sur un même espace (Ω,α) on définit la distance $K(F,G) = \sup_{A\in\mathcal{A}} |F(A)-G(A)|$, puis la classe

$$\mathcal{F}_K(\varepsilon) = \{G \mid \exists\ \theta\in\Theta \text{ tel que } K(F_\theta,G) \leq \varepsilon\}$$

qui est la réunion quand θ varie dans Θ des boules de centre F_θ et de rayon ε.

Remarque : si $H = (1-\varepsilon)F + \varepsilon G$ on a :

$$K(F,H) = \sup_{A\in\mathcal{A}} |(1-\varepsilon)F(A) + \varepsilon G(A) - F(A)| = \sup_{A\in\mathcal{A}} \varepsilon| F(A)-G(A)| \leq \varepsilon.$$

33 - Modèle de Prokhorov.

$$\mathcal{F}_\pi(\varepsilon) = \{G \mid \exists\ \theta\in\Theta \text{ tel que } \pi(F_\theta,G) \leq \varepsilon\}$$

Ce modèle est très utilisé : il permet de tenir compte non seulement de la non-équidistribution mais également de la non-indépendance. Hampel [7] a bâti sur ce modèle une mesure qualitative de robustesse.

A présent de nombreux estimateurs robustes ont été recensés et il est parfois difficile de savoir lequel choisir.

§ II - 3 SIMILARITE DES ESTIMATEURS

Introduction : Hampel [8] indique que plus de 70 estimateurs de position différents ont été recensés et étudiés du point de vue de la robustesse dans le cadre de l'étude de Princeton.

Une question se pose alors : pourquoi utiliser tel estimateur plutôt que tel autre ? Quelles sont leurs différences ou leurs points communs ? Il serait donc utile de dégager certaines familles ou de mettre en évidence certaines proximités. Mais on ne dispose pas d'une distance entre estimateurs. On a donc essayé de trouver un moyen pour dégager un ou deux facteurs qui sont de bons séparateurs de façon à réaliser une "carte des estimateurs". On présente, ici, le début de cette étude.

Dans un premier temps on a repéré les échantillons que l'on a trouvés dans [16] . On pourra par la suite y ajouter d'autres échantillons, voire en produire un grand nombre par simulation - ce qui permettrait de préciser les conclusions obtenues pour un modèle particulier : celui de Prokhorov pour des lois gaussiennes par exemple -, et surtout on pourra augmenter considérablement le nombre des estimateurs de façon à obtenir une carte aussi complète que possible. Cette étude demande un gros travail de programation et l'on espère la poursuivre et la compléter ultérieurement.

34 - <u>Les estimateurs</u> :

Nous décrivons brièvement les estimateurs étudiés.

Soit $x_1,\dots, x_n$ un n échantillon. On note $x_{(1)},\dots,x_{(n)}$ l'échantillon ordonné dans le sens croissant.

E_1 la moyenne arithmétique

$$\text{MOYE} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

E_2 la médiane

$$\text{média} = \begin{cases} x_{(p+1)} & \text{si } n = 2p+1 \\ \frac{1}{2}[x_{(p)}+x_{(p+1)}] & \text{si } n = 2p \end{cases}$$

Les moyennes α-tronquées : si α est un pourcentage

La moyenne α-tronquée est définie par :

$$\bar{x}_\alpha = \frac{1}{n(1-2\alpha)} \left[c(x_{([\alpha n+1])} + x_{(n-[\alpha n])}) + \sum_{i=[\alpha n+2]}^{n-[\alpha n+1]} x_{(i)}\right]$$

où $c = 1+[\alpha n] - \alpha n$

E_5 : moyenne 5 % tronquée $1 - 2\alpha = 0,9$

E_6 : moyenne 15 % tronquée $1 - 2\alpha = 0,7$

E_7 ; moyenne 25 % tronquée $1 - 2\alpha = 0,5$

Les M-estimateurs

Ce sont des estimateurs introduits par Huber et définis implicitement comme une solution de l'équation

$$\sum_{i=1}^{n} \phi \frac{(x_i - T_n)}{s} = 0$$

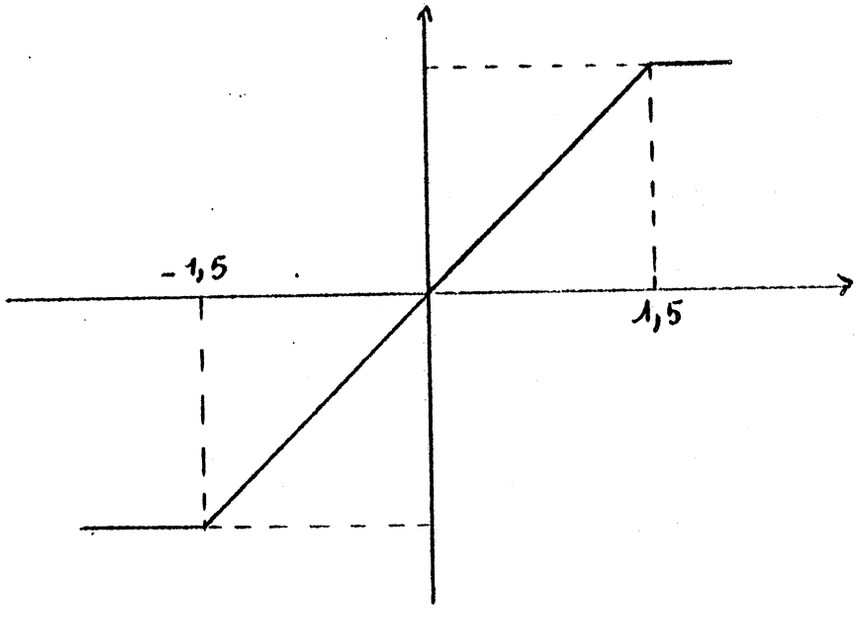
où T_n est le M-estimateur, x_i les observations et s un estimé de la dispersion que l'on précisera ; ϕ est une fonction impaire. On sait que sous certaines conditions sur la fonction ϕ les M-estimateurs sont robustes et la fonction d'influence est donnée par (Andrews et al [3])

$$IC(\xi) = \frac{s\ \phi\ (\frac{\xi}{s})}{\int \phi'(\frac{y}{s})\ d\ F(y)}$$

calcul : il faudra préciser s et le nombre d'itérations pour le calcul approché de la solution.

E_8 : M estimateur de Huber.

Soit $\phi(u) = \min(1,5\ ;\ \max(-1,5,u))$



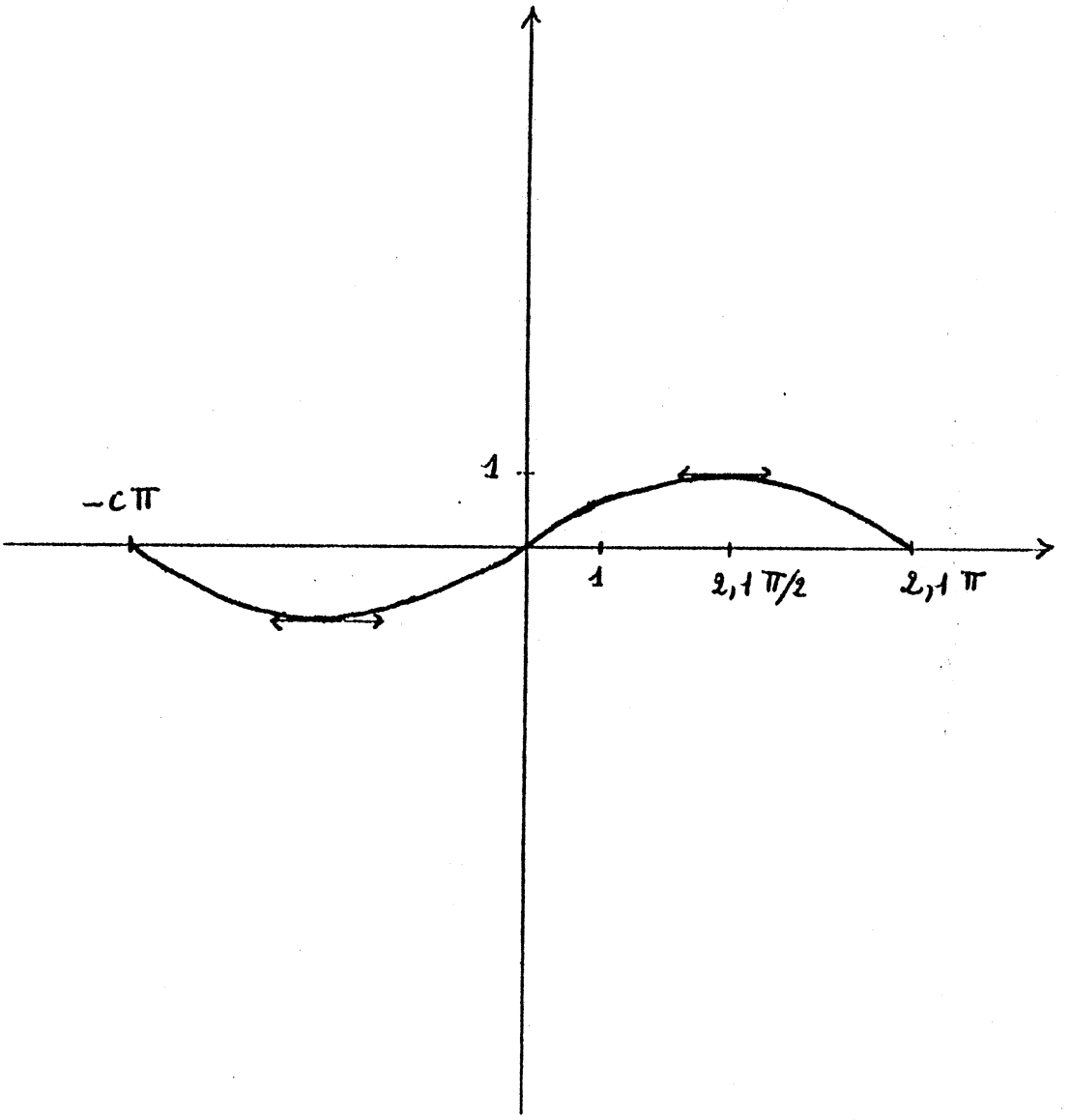
nous ne ferons qu'une seule itération et pour s nous prendrond l'écart absolu médian autour de la médiane.

E_9 <u>M-estimateur de Andrews</u>

La fonction ϕ est déterminée par

$$\phi(u) = \begin{cases} \sin(\frac{u}{2.1}) & \text{si} \quad |u| < (2,1)\pi \\ 0 & \text{sinon} \end{cases}$$

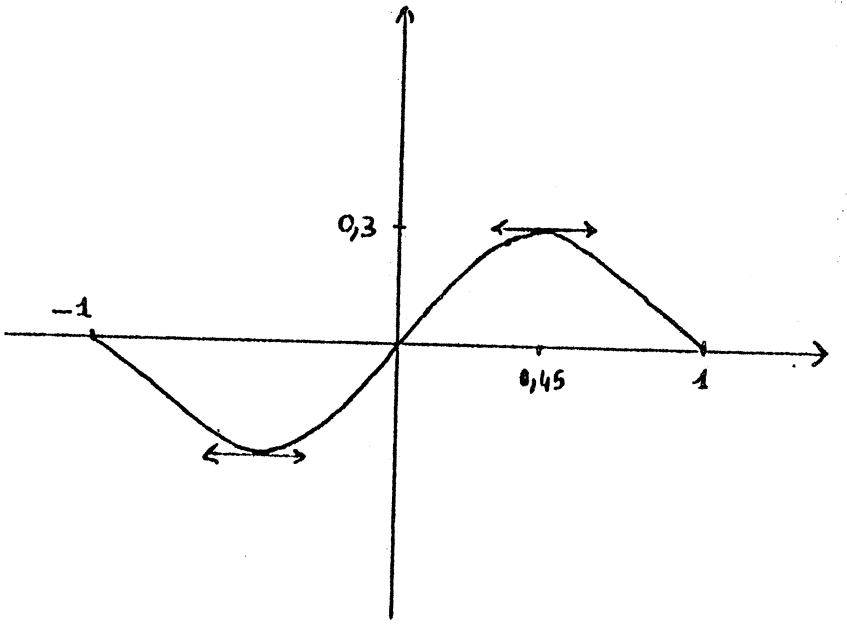
où s est l'écart absolu médian à une valeur de T_n obtenue lors du calcul, en commençant par la médiane et changée toutes les 3èmes itérations.



E_{10} <u>M-estimateur de Tukey</u> (encore appelé bi-poids)

La fonction ϕ est définie par

$$\phi(u) = \begin{cases} u(1-u^2)^2 & \text{si} \quad |u| \leq 1 \\ 0 & \text{si} \quad |u| > 1 \end{cases}$$



calculé par itérations (six en commençant par s = écart absolu médian à la médiane)

E_3 <u>Estimateur de Edgeworth</u>

Moyenne pondérée des quartiles avec des poids proportionnels à 5, 6, 5

$$E_3 = \frac{1}{16}(5h_1 + 6\text{Med} + 5h_2)$$

h_1 et h_2 valeurs approchées des quartiles empiriques (voir Andrews et al [3])

E_4 <u>Moyenne décalée</u>

Définie par

$$\bar{x}^c_{.25} = 2\,E_1 - E_5$$

E_{11} Estimateur adaptatif de Hogg

Soit

$$Q = \frac{U(0,05) - L(0,05)}{U(0,5) - L(0,5)}$$

où $L(\alpha)$ et $U(\alpha)$ sont les moyennes des $100x\alpha\%$ plus petits (resp-plus grands) des x_i.

Alors
$$E_{11} = \begin{cases} E_7 & \text{si } Q \quad 2,0 \\ E_1 & \text{si } 2,0 \quad Q \quad 2,6 \\ \bar{x}_{3/16} & \text{si } 2,6 \quad Q \quad 3,2 \\ \bar{x}_{3/8} & \text{si } 3,2 \quad Q \end{cases}$$

Il est clair qu'à cette liste on pourrait en ajouter bien d'autre : les M-estimateurs de Hampel [10] (de type th, ou Arctang), les estimateurs du type test + estimation, par ex. : test de Grubbs , Thompson ou Shapiro-Wilh [9] , la médiane pliée de Hodges-Lehmann [3] et bien d'autres.

35 - Les données

On trouvera les données dans [16] . Il s'agit de mesures historiques effectuées en astronomie :

- les déterminations faites par James Short en 1763 de la parallaxe du soleil à partir du transit de Vénus de 1761

- les déterminations de la densité de la terre faite par Cavendish en 1798

- les déterminations de la vitesse de la lumière par Michelson et Newcomb en 1879 et 1882.

Stigler en a fait 24 échantillons que l'on reprend ici. Ces échantillons ont été centrés et réduits. Puis on a fait opérer les onze estimateurs sur ces 24 nouveaux échantillons. Les résultats sont consignés dans le tableau ci-dessous. On a alors effectué une analyse en composantes principales normée [15] , en condi-

```
es 1 0.0000011 -0.0000006 -0.0000011  0.0000005 -0.0000005 -0.0000010  0.0000005  0.0000014 -0.0000025  0.0000000 -0.0000004  0.0000
005  0.0000000  0.0000000  0.0000005 -0.0000005 -0.0000004 -0.0000004 -0.0000041  0.0000003 -0.0000002 -0.0000003  0.0000003  0.0000
000
es 2-0.1814400 -0.0202500 -0.3220700 -0.0822600 -0.3386700 -0.1456700 -0.0173400 -0.0137600  0.2127350 -0.1073300 -0.1345300  0.2954
450 -0.1798450  0.1264150 -0.0916050 -0.3965400  0.1660100 -0.1233000  0.0546200 -0.0600500 -0.1551700 -0.1534200  0.0733200 -0.0303
800
es 3-0.0823963  0.0550469 -0.2142950 -0.1728912 -0.2325950 -0.1930513  0.0281725  0.0058650  0.2127350 -0.0463456 -0.0482425  0.1465
325 -0.1798434  0.1659150 -0.0655784 -0.0795384  0.0405631 -0.0412433  0.0404762 -0.0298156 -0.0049731 -0.0810262  0.1024044 -0.0303
769
es 4 0.0954444 -0.0265200  0.1968700  0.1329590  0.2287105  0.1308448 -0.0059438  0.0386033 -0.1315390  0.0663500  0.0364808 -0.2077
650  0.0784770 -0.1137710  0.0066630  0.1696310 -0.0857339  0.0631365 -0.0396138  0.0171414  0.1005419  0.1409644 -0.0936619  0.0418
178
es 5-0.0510389  0.0388566 -0.1409132 -0.1147298 -0.1667232 -0.1225792 -0.0241432 -0.0425560  0.1896887 -0.0035381 -0.0306558  0.1393
850 -0.0653975  0.1422125 -0.0072856 -0.0818438  0.0011685 -0.0425636  0.0482966 -0.0115754 -0.0223229 -0.0726952  0.1074636 -0.0043
033
es 6-0.0656214  0.0535950 -0.1907317 -0.1525233 -0.2082381 -0.1359503 -0.0153320 -0.0407211  0.1904436 -0.0501800 -0.0387231  0.1647
429 -0.1374386  0.1354407 -0.0190336 -0.0895835  0.0485565 -0.0496366  0.0399744 -0.0276094 -0.0428372 -0.0988710  0.1072500 -0.0150
434
es 7-0.0954422  0.0265183 -0.1968722 -0.1329581 -0.2287114 -0.1308467  0.0059448 -0.0386010  0.1815340 -0.0663500 -0.0364815  0.2077
660 -0.0784770  0.1137710 -0.0066620 -0.1696820  0.0857330 -0.0631374  0.0396055 -0.0171407 -0.1005423 -0.1409651  0.0936624 -0.0418
178
es 8-0.0445610  0.0509146 -0.1392397 -0.1194246 -0.1283187 -0.1039626 -0.0331500 -0.0460462  0.2078072 -0.0041543 -0.0239838  0.0915
089 -0.0010121  0.1625286  0.0000005 -0.0646408  0.0169243  0.0003033  0.0306328  0.0057384 -0.0366367 -0.0893048  0.1105349 -0.0013
017
es 9-0.0455782  0.0064525 -0.1565116 -0.1525226 -0.1630295 -0.1230703 -0.0362469 -0.0660079  0.2741127 -0.0135313 -0.0152330  0.0896
663 -0.0198704  0.2574421 -0.0019005 -0.0702728 -0.0325458 -0.0104770  0.0330217 -0.0030553 -0.0770140 -0.1010081  0.1367266 -0.0019
858
es10-0.0766775 -0.2694696 -0.2187855 -0.2489743 -0.2492556 -0.2209107 -0.0437567 -0.0566818 -1.2123189 -0.0272586 -0.0562228  0.1703
499 -0.0396107 -0.1457063 -0.0060592 -0.1579230  0.0175312 -0.0200479  0.0482779 -0.0080340 -0.3211355 -0.1674321 -2.3900812 -0.0095
221
es11 0.0954444 -0.0265200  0.1968700  0.0000005 -0.2332267 -0.1522644 -0.0171718 -0.0403367  0.1701880  0.0000000 -0.1032615  0.1845
120  0.0784770  0.1011320  0.0066630 -0.1003336  0.0637579 -0.0000004  0.0387443  0.0000003 -0.1260468 -0.1179134  0.0936624 -0.0281
078
```

dérant le système : variables-valeurs des estimateurs pour les échantillons, individus-estimateurs. Pour les calculs sur ordinateur on a utilisé des sous-programmes des bibliothèques S.P.S.S., SPAD ainsi que certains programmes qui sont décrits dans [3] . On résume ici les résultats de cette analyse.

On fournit l'histogramme des valeurs propres, la projection des individus sur le premier facteur et dans le plan des deux premiers facteurs.

Ce qui frappe le plus c'est que 3 à 4 facteurs portent la quasi-totalité de l'inertie. Le premier facteur en assumant près de 60 % et les deux premiers près de 80 %. On constate alors que les estimateurs s'échelonnent le long du premier axe en fonction de l'importance (1) qu'ils accordent aux valeurs extrêmes.

La moyenne décalée qui accroit le poids de ces valeurs est à l'extrémité droite. Vient ensuite la moyenne arithmétique, puis les moyennes α-tronquées dans l'ordre d'importance de troncature et les M-estimateurs qui se comportent un peu comme des estimateurs à winsorisation. La médiane qui ne tient compte que "du milieu" de l'échantillon est à l'extrémité gauche. On peut éventuellement dégager un groupe à partir des M-estimateurs E_8-E_9-E_{10} de la moyenne 10 % tronquée et de l'estimateur de Hogg.

Il se donfirme donc que le traitement des valeurs extrêmes est une étape fondamentale de l'estimation.

(1) Notons que la valeur de la projection des estimateurs sur le premier facteur peut nous fournir une mesure de cette importance, lorsqu'on peut difficilement l'évaluer, comme pour les M-estimateurs.

EDITION DES VALEURS-PROPRES

OMME DES VALEURS-PROPRES 24.00000024

ISTOGRAMME DES PREMIERES VALEURS-PROPRES

	VALEUR-PROPRE	POURCENTAGE	POURCENTAGE CUMULE	
1	13.85618603	57.73	57.73	**********
2	4.85739877	20.24	77.98	**********
3	2.39357826	9.97	87.95	**********
4	1.45304090	6.05	94.00	********
5	0.87059403	3.63	97.63	*****
6	0.32596408	1.36	98.99	**
7	0.12548417	0.52	99.51	*
8	0.07076781	0.29	99.81	*
9	0.04648605	0.19	100.00	*
10	0.00000007	0.00	100.00	*
11	0.00000003	0.00	100.00	*
12	0.00000002	0.00	100.00	*
13	0.00000001	0.00	100.00	*
14	0.00000001	0.00	100.00	*
15	0.00000000	0.00	100.00	*

ITION SOMMAIRE DES VALEURS-PROPRES DE 16 A 24

).00000000 0.00000001 0.00000001 0.00000002 0.00000003 0.00000003 0.00000005 0.00000008 0.00000009

AIDE A L INTERPRETATION DES AXES — PROJECTION DES POINTS SUR L-AXE 1

ec12	es 2
ec19	
ec17	
ec14	es 7
	es 3
	es10
ec09	es 6
ec02	es 5
ec07	es 9
ec23	
	es11
ec08	es 8
ec21	
ec15	
ec04	es 1
ec13	
ec03	
ec20	
ec01	
ec06	
ec24	
ec11	
ec22	
ec18	
ec16	
ec10	
ec05	
	es 4

1 GRAPHIQUES DANS LE GROUPE 3

LIGNES ACTIVES = 11 ILLUSTRATIVES = 0

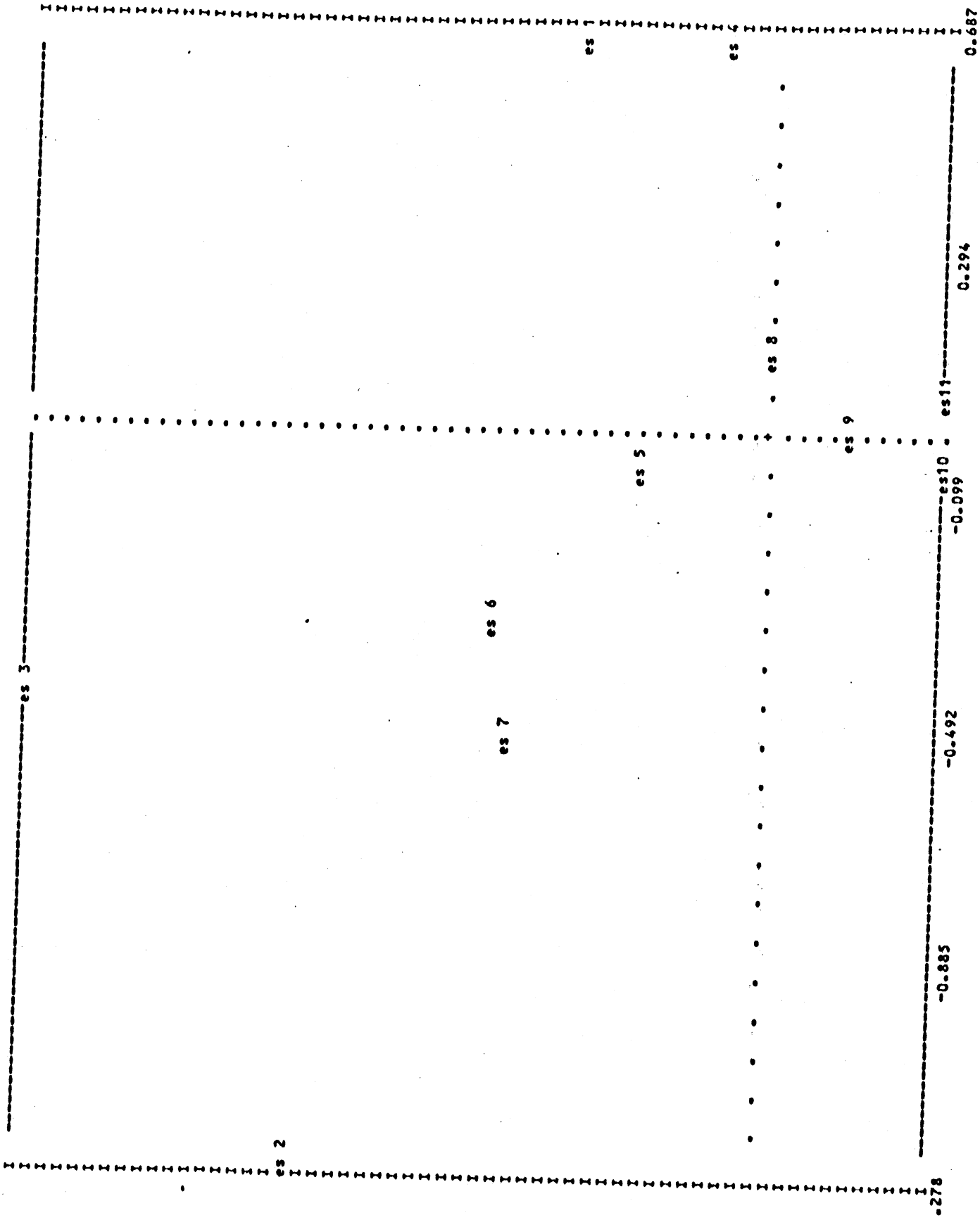
PLAN DE PROJECTION DES 11 POINTS SUR LES AXES 1 ET 2

AXE 1 /HORIZONTAL AXE 2 /VERTICAL

ATTENTION

LES POINTS CI-DESSOUS ETAIENT A PLUS DE 2.3 ECARTS-TYPES DU CENTRE
ILS ONT ETE RAMENES SUR LE CADRE DU GRAPHIQUE

es 4	1.88087	0.08344
es10	-0.38845	-1.32637


es 1
es 4
0.687
0.294
es 8
es11
es 9
es 5
es10
-0.099
es 6
es 3
es 7
-0.492
-0.885
es 2
.278

BIBLIOGRAPHIE DU CHAPITRE II

[1] ANDREWS D.F., 1971 : Plotting high dimensional data. 38ème Session I.S.I. Washington D.C., P. 22-26

[2] ANDREWS D.F., 1972 : Plots of high dimensional data. Biometrics 28,P. 125-36

[3] ANDREWS D.F. et CIE, 1972 : Robust estimates of location P.U.P. New Jersey.

[4] BARNETT V. et LEWIS T., 1978 : Outliers in statistical data. Wiley and Sons, New-York.

[5] EVERITT B.S. et NICHOLLS P., 1975 : Visual techniques for representing multivariate Data. The statistician, vol 24, n°1, P. 37-49.

[6] GAREL B., 1976 : Détection des valeurs aberrantes dans un échantillon gaussien multidimensionnel. Thèse de 3ème cycle Grenoble.

[7] HAMPEL F.R., 1971 : A general qualitative definition of robustness. A.M.S. vol 42 n°6 P. 1887-96.

[8] HAMPEL F.R., 1973 : Robust estimation. A condensed partial survey. Whars.verw. Gerb. 27, P. 87-104 (Springer-Verlag).

[9] HAMPEL F.R., 1976 : On the rejection point of some rejection rules with mean. Research Report n°11 E.T.H., Zuerich.

[10] HAMPEL F.R., ROUSSEEUV P.J., RONCHETTI E. : The change of variance curve and optimal descending M-estimators. Research Report n°22, E.T.H. Zuerich, 1980.

[11] HUBER P.J., 1964 : Robust estimation of a location parameter. A.M.S. 35 P. 73-101.

[12] HUBER P.J., 1981 : Robust Statistics ; Wiley.

[13] QUITTARD-PINON F., 1973 : Quelques méthodes d'analyse des données aberrantes ; thèse de 3ème cycle, Paris.

[14] ROHLF F.J.: generalisation of the gap test for the detection of multivariate outliers. Biometrics 31, P. 93-101

[15] SAPORTA G., 1978 : Théories et méthodes de la statistique ; Editions Technip.

[16] STIGLER S.M., 1977 : Do robust estimators work with real data ? Annals of Stat, vol 5, n°6, PP. 1055-1099.

[17] TUKEY J.W., 1960 : A survey of sampling from contaminated distributions. In OLKIN P. 448-485 : Contributions to probability and statistics. Stanford U. Press.

[18] VAN CUSTEM B. (1972) : Un théorème de convergence des suites d'estimateurs ensemblistes du maximum de vraisemblance ; Grenoble IMAG.

[19] GNEDENKO, 1943 : sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math 44 - P 423-453.

[20] DAVID, 1970 : Order Statistics, Wiley

CHAPITRE III

SEPARATION ASYMPTOTIQUE DES CHAINES DE MARKOV

Résumé :

Soit (X_n) et (Y_n) deux suites de variables aléatoires à valeurs dans un ensemble fini de cardinal m. On suppose que (X_n) (resp. (Y_n)) constitue une chaîne de Markov et qu'elle admet P (resp. Q) comme matrice de transition stationnaire.

On trouve une condition nécessaire et suffisante de non séparation asymptotique.

§ III - 1 Introduction :

On trouvera dans Hillion [9] et [10] une étude détaillée des notions que l'on rappelle ici. Soit $(\mathcal{X}, \mathcal{B})$ un espace mesurable. Soient P et Q deux probabilités définies sur $\mathcal{B}$.

Mesure sur $(\mathcal{X}, \mathcal{B})$ et distances entre probabilités.

On définit la distance de Hellinger entre P et Q

(1) $$d_H^2(P,Q) = \frac{1}{2}\int (\sqrt{dP} - \sqrt{dQ})^2 = 1 - \int \sqrt{dP.dQ}$$

et l'on note $\rho(P,Q) = \int \sqrt{dP.dQ}$ l'affinité de Hellinger. On peut également définir la distance en variation totale entre P et Q par :

(2) $$d_V(P,Q) = \sup_{B \in \mathcal{B}} |P(B) - Q(B)|$$

Ces deux distances sont liées par les inégalités (cf. Moché [12])

(3) $$d_H^2 \leq d_V \leq \sqrt{2}\, d_H$$

Supposons maitenant que les probabilités P et Q soient dominées par une mesure μ, σ-finie. Appelons f (resp. g) la densité de P (resp. Q) par rapport à μ. La distance de Hellinger s'écrit alors :

$$(4) \qquad d_H^2(P,Q) = \frac{1}{2}\int (\sqrt{f} - \sqrt{g})^2 \, d\mu$$

Soit $\alpha \in]0, 1[$. On peut définir sur $\mathcal{B}$ la mesure $P^\alpha Q^{1-\alpha}$ par :

$$\forall B \in \mathcal{B} \qquad P^\alpha Q^{1-\alpha}(B) = \int_B f^\alpha g^{1-\alpha} \, d\mu$$

Si l'on note $H_\alpha(P,Q) = P^\alpha Q^{1-\alpha}(\mathcal{X})$ on a :

$$(5) \qquad H_{1/2}(P,Q) = \rho(P,Q)$$

<u>Orthogonalité de deux probabilités</u> :

On dit que les deux probabilités P et Q sont orthogonales et l'on note $P \perp Q$ si et seulement si $\exists\ B \in \mathcal{B}$ tel que $P(B) = Q(B^c) = 0$.

Cette relation est alors successivement équivalente à :

i) $d_V(P,Q) = 1$ ii) $d_H(P,Q) = 1$ iii) $\rho(P,Q) = 0$

§ III - 2 <u>Analyse spectrale du problème</u>.

Nous mettons maintenant en place le cadre dans lequel nous allons poursuivre l'étude du problème. Nous nous inspirons des méthodes algébriques dans les chaînes de Markov utilisées par Karlin [8] , Pallu de la Barrière [13] . On pourra également consulter Kolmogorov et Fomine [11] sur les fonctionnelles et opérateurs linéaires.

Soit $x \in \mathbb{R}^m$. On utilise la norme :

$$(6) \qquad ||x|| = \sum_{i=1}^{m} |x_i|$$

Soit $L(\mathbb{R}^m, \mathbb{R}^m)$ l'ensemble des matrices carrées mxm formées de nombres réels.

Pour $A \in L(\mathbb{R}^m, \mathbb{R}^m)$ on notera A' la matrice transposée de A. On définit la norme suivante :

(7) $$||A|| = \sup_{||x|| \leq 1} ||A'x||$$

On notera $A = (a_{ij})$ une matrice à m lignes et m colonnes d'éléments a_{ij} où le premier indice numérote les lignes et le deuxième indice numérote les colonnes. On pose :

$$A_{i.} = \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{im} \end{pmatrix}$$

8 - <u>Lemme</u> : <u>On a la relation</u> :

$$||A|| = \max_{i=1,\dots,m} ||A_{i.}||$$

Preuve : soit $x \in \mathbb{R}^m$

$$A'x = (\sum_{i=1}^{m} a_{ij}\, x_i) = \sum_{i=1}^{m} x_i A_{i.}$$

Alors d'après l'inégalité triangulaire :

$$||A'x|| = ||\sum_{i=1}^{m} x_i\, A_{i.}|| \leq \sum_{i=1}^{m} |x_i|\, ||A_{i.}|| \leq \max_{i=1,\dots,m} ||A_{i.}|| \cdot \sum_{i=1}^{m} |x_i|$$

$$= \max_{i=1,\dots,m} ||A_{i.}||\, ||x||$$

D'où l'on tire $||A|| \leq \max_{i=1,\dots,m} ||A_{i.}||$

Inversement, supposons que $\max_{i=1,\dots,m} ||A_{i.}|| = ||A_{i_0.}||$

Soit e_{i_0} le $i_0$ème vecteur de base de $\mathbb{R}^m$, i.e. le vecteur dont la ième composante est égale à 1 et toutes les autres nulles :

$$e_{i_o} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Alors $||e_{i_o}|| = 1$ et $A'e_{i_o} = A_{i_o}$

Donc $||A|| \geq ||A'e_{i_o}|| = \max\limits_{i=1,\dots,m} ||A_{i_o}||$. D'où le résultat.

Soient α,β des entiers tels que $\alpha+\beta = m$. On notera $A_{\beta\alpha}$ un élément de $L(\mathbb{R}^\beta, \mathbb{R}^\alpha)$.

Pour $\alpha < m$ on pourra considérer que $A_{\alpha\alpha} \in L(\mathbb{R}^m, \mathbb{R}^m)$ en identifiant

$$A_{\alpha\alpha} \text{ avec } \begin{pmatrix} A_{\alpha\alpha} & 0 \\ 0 & 0 \end{pmatrix}$$

les 0 constituant des matrices blocs dont les éléments sont nuls.

9 - <u>Corollaire</u>. <u>Supposons que la matrice $A \in L(\mathbb{R}^m, \mathbb{R}^m)$ se décompose sous la forme de matrice-blocs de la façon suivante</u> :

$$A = \begin{pmatrix} A_{\alpha\alpha} & 0 \\ A_{\beta\alpha} & A_{\beta\beta} \end{pmatrix}$$

<u>Alors</u> $||A|| \geq ||A_{\alpha\alpha}||$

Preuve : d'après le lemme 8, $||A|| = \max\limits_{i=1,\dots,m} ||A_{i.}|| \geq \max\limits_{j=1,\dots,\alpha} ||A_{j.}|| = ||A_{\alpha\alpha}||$

10 - <u>Lemme</u> : <u>Soient A et B deux matrices</u> $\in L(\mathbb{R}^m, \mathbb{R}^m)$

<u>alors</u> $||A.B|| \leq ||A||.||B||$.

Preuve : posons $A = (a_{ij})$ et $B = (b_{ik})$

i, j, k variant de 1 à m.

$$A.B = \left(\sum_{j=1}^{m} a_{ij}\, b_{jk}\right) . \text{ On a } (A.B)_{i.} = \begin{pmatrix} \sum_{j=1}^{m} a_{ij}\, b_{j1} \\ \vdots \\ \sum_{j=1}^{m} a_{ij}\, b_{jm} \end{pmatrix}$$

$$||(A.B)_i|| = \sum_{k=1}^{m} \left| \sum_{j=1}^{m} a_{ij}\, b_{jk} \right| \leq \sum_{k=1}^{m} \sum_{j=1}^{m} |a_{ij}\, b_{jk}|$$

$$= \sum_{k=1}^{m} \sum_{j=1}^{m} |a_{ij}|\,|b_{jk}| = \sum_{j=1}^{m} |a_{ij}| \sum_{k=1}^{m} |b_{jk}| = \sum_{j=1}^{m} |a_{ij}|\; ||B_{j.}||$$

$$\leq \max_{j=1,\dots,m} ||B_{j.}|| \sum_{j=1}^{m} |a_{ij}| \leq \max_{j=1,\dots,m} ||B_{j.}|| \max_{i=1,\dots,m} ||A_{i.}||$$

$$= ||B||.||A||$$

11 - <u>Lemme</u> : <u>Soit</u> $A \in L(\mathbb{R}^m, \mathbb{R}^m)$ <u>et</u> $x \in \mathbb{R}^m$.

<u>Alors</u> $(||A'x|| = ||A||\;||x||) \Rightarrow (\forall i$ <u>tel que</u> $x_i \neq 0\; ||A_{i.}|| = ||A||)$

Preuve : on a alors $||A'x|| = \sum_i |x_i|\; ||A_{i.}|| = \max_i ||A_{i.}|| \sum_{i=1}^{m} |x_i| = ||A||\;||x||$.

La seconde égalité ne peut avoir lieu que si $||A_{i.}|| = \max_{i=1,\dots,m} ||A_{i.}||$

§ III - 3 <u>Séparation aymptotique</u>.

12 - <u>Définition</u>. <u>On appelle matrice sous-markovienne</u> (<u>resp. markovienne ou stochastique ou matrice de transition</u>) <u>une matrice</u> $A = (a_{ij})$ <u>telle que</u> $\forall\, i,j\;\; a_{ij} \geq 0$ <u>et</u> $\forall\, i\;\; \sum_j a_{ij} \leq 1$ (resp. $\sum_j a_{ij} = 1$).

Le produit de deux matrices sous-markoviennes (resp. stochastiques) est une matrice sous-markovienne (res.stochastique). Une matrice sous-markovienne (resp. stochastique) est de norme ≤ 1 (resp.1). Une matrice stochastique admet la valeur propre 1. Soient (X_n) et (Y_n) deux suites variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(\mathcal{X}, \mathcal{B})$ où $\mathcal{X} = \{1,\dots,m\}$ et constituant chacune une chaîne de Markov homogène de même loi initiale et de matrices de

transition respectives P et $Q \in L(\mathbb{R}^m, \mathbb{R}^m)$.

Le vecteur $(X_1, \ldots, X_n)$ est alors à valeurs dans $(\mathcal{X}^{(n)}, \mathcal{B}^{(n)}) = \bigotimes_{i=1}^{n} (\mathcal{X}_i, \mathcal{B}_i)$ où pour tout i: $(\mathcal{X}_i, \mathcal{B}_i) = (\mathcal{X}, \mathcal{B})$.
On notera $P^{(n)}$ la loi conjointe du vecteur $(X_1, \ldots, X_n)$ et $Q^{(n)}$ celle du vecteur $(Y_1, \ldots, Y_n)$

13 - <u>Définition</u> : <u>on dira qu'il y a séparation asymptotique des deux chaînes de Markov</u> (X_n) <u>et</u> (Y_n) <u>si et seulement si</u> $\lim_{n \to +\infty} d_V (P^{(n)}, Q^{(n)}) = 1$.

Cette définition est alors équivalente à (voir Hillion [10])

$$\lim_{n \to +\infty} d_H (P^{(n)}, Q^{(n)}) = 1 \text{ ou encore à}$$

$$\lim_{n \to +\infty} \rho (P^{(n)}, Q^{(n)}) = 0.$$

Par définition des matrices $P = (p_{ij})$ et $Q = (q_{ij})$
on a : $\forall\, i, j \in \{1, \ldots, m\}$ $p_{ij} = \mathbb{P}(x_n = j \,/\, x_{n-1} = i)$ $q_{ij} = \mathbb{P}(y_n = j \,/\, y_{n-1} = i)$.
On définit alors la matrice $\mathcal{R}$ par $\mathcal{R} = (\rho_{ij})$ où $\rho_{ij} = (p_{ij})^{1/2} (q_{ij})^{1/2}$

14 - <u>Lemme</u>. <u>La matrice $\mathcal{R}$ est sous markovienne</u>. <u>On aura</u> $||\mathcal{R}|| = 1$ <u>si et seulement si</u> $\exists\ i_0 \in \{1, \ldots, m\}$ <u>tel que</u> $\forall\, j = 1, \ldots, m$ $p_{i_0 j} = q_{i_0 j}$

Preuve : la somme des termes de la ième ligne de la matrice $\mathcal{R}$ nous est donnée par $\sum_{j=1}^{m} \sqrt{p_{ij} q_{ij}}$. Appliquons l'inégalité de Cauchy-Schwarz aux suites $(\sqrt{p_{ij}})$, $(\sqrt{q_{ij}})$.

Il vient : $\left(\sum_{j=1}^{m} \sqrt{p_{ij}} \sqrt{q_{ij}}\right) \leq \left(\sum_{j=1}^{m} p_{ij}\right)^{1/2} \left(\sum_{j=1}^{m} q_{ij}\right)^{1/2} = 1.$

Pour que $||\mathcal{R}|| = 1$ il faut qu'$\exists\ i_0$ tel que $||\mathcal{R}_{i_0 \cdot}|| = 1$.
On sait qu'il n'y a égalité dans la formule de Cauchy-Schwarz que si $\exists\ \lambda \in \mathbb{R}$ tel que $P_{i_0 \cdot} = \lambda\, Q_{i_0 \cdot}$. On en déduit que $\sum_{j=1}^{m} p_{i_0 j} = \lambda \sum_{j=1}^{m} q_{i_0 j}$ soit $\lambda = 1$ et

$p_{i_0 j} = q_{i_0 j} \quad \forall\ j=1,\dots,m.$

On en déduit la seconde partie du lemme.

15 - <u>Proposition</u> : <u>Une matrice sous-markovienne A qui admet la valeur propre 1 contient une sous-matrice stochastique</u>.

Si 1 est valeur propre de A on sait que la dimension de la variété linéaire des vecteurs propres à gauche de A est égale à l'ordre de multiplicité de la valeur propre A donc au moins égale à 1. Soit x un tel vecteur.
Supposons que x ait au moins une composante nulle. Alors en réindiçant si nécessaire l'espace des états on peut supposer que

$$x = \begin{pmatrix} x_{i_1} \\ \vdots \\ x_{i_k} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} x_{(k)} \\ 0 \end{pmatrix}$$

avec $\forall\ l = 1,\dots,k \qquad x_{i_l} \neq 0.$

En effet, effectuer une permutation entre deux indices i et j de l'ensembles des états revient à permuter la ligne i avec la ligne j puis la colonne i et la colonne j dans la matrice de transition. Ceci entraine alors une permutation des composantes i et j d'un vecteur propre.

Ecrivons A sous la forme :

$$A = \begin{pmatrix} A_{kk} & A_{k,m-k} \\ A_{m-k,k} & A_{m-k,m-k} \end{pmatrix}$$

Par hypothèse on a $x'A = x'$ et l'on voit que $x'_{(k)}A_{kk} = x'_{(k)}$.

D'où $||x'_{(k)}A_{kk}|| = ||A'_{kk}x_{(k)}|| = ||A_{kk}||\ ||x_{(k)}|| = ||x_{(k)}||$.
Ce qui entraine que $||A_{kk}|| = 1$.

Alors, d'après le lemme 11, toutes les lignes de la matrice A_{kk} sont de norme 1. Par suite $A_{k,m-k} = 0$ et A s'écrit :

$$A = \begin{pmatrix} A_{kk} & 0 \\ A_{m-k,k} & A_{m-k,m-k} \end{pmatrix}$$

où A_{kk} est une matrice stochastique.

Enfin si toutes les composantes de x sont $\neq 0$ on a alors $k = m$ et $A_{kk} = A$.

Nous rapellons maintenant quelques définitions concernant les états d'une chaîne de Markov. On pourra consulter K.L. Chung [4] , Kemeny et Snell [7] ou Revuz [14]. Pour tout $i \in \mathcal{X}$ notons T_i le temps d'arrêt $\{\inf n > 0 \ / \ X_n = i\}$. Un état est récurrent si $\mathbb{P}(T_i < +\infty) = 1$; il est transitoire (transient) dans le cas contraire.

Si l'on note $\mathbb{P}_i(.) = \mathbb{P}(./X_1 = i)$

posons $\rho_{ij} = \mathbb{P}_i(T_j < +\infty)$.

On dira que deux états i et j communiquent si $\rho_{ij}.\rho_{ji}$ 0. C'est une relation d'équivalence. Dans notre cas ($\mathcal{X}$ fini) une chaîne de Markov a au moins un état récurrent, donc une classe récurrente.

16 - <u>Définition</u> : <u>Un sous ensemble $\mathcal{C}$ de $\mathcal{X}$ est clos si pour tout i $\in \mathcal{C}$</u> $\mathbb{P}_i(X_2 \in \mathcal{C}) = 1$.

On voit facilement qu'une condition nécessaire et suffisante pour que $\mathcal{C}$ de cardinal k soit clos est que la matrice de transition, après une réindexation éventuelle des états, s'écrive :

$$A_{\cdot} = \begin{pmatrix} A_{kk} & 0 \\ A_{m-k,k} & A_{m-k,m-k} \end{pmatrix}$$

On peut alors énoncer le résultat suivant

17 - <u>Théorème</u> : <u>Les assertions suivantes équivalentes</u>

1°) $\lim_{n \to +\infty} \rho(P^{(n)}, Q^{(n)}) = 0$

2°) $\lim_{n \to +\infty} \mathcal{R}^n = 0$

3°) $\det (I - \mathcal{R}) \neq 0$

Ce résultat est connu de Darwich [6]. Mais comme l'introduction du produit décomposé simplifie considérablement la preuve, nous en donnons ici une nouvelle version. On trouvera en appendice une introduction sommaire et quelques propriétés du produit décomposé.

<u>Preuve</u> : appelons L_0 le vecteur décrivant la loi initiale.

$$\rho(P^{(n)}, Q^{(n)}) = P^{(n)^{1/2}} Q^{(n)^{1/2}} (\mathcal{X}^{(n)})$$

Les élements de $\mathcal{X}^{(n)}$ étant des atomes.

Alors on sait que la probabilité $P^{(n)}$ est décrite par

$$L_0 \chi(\underbrace{P\chi(P\chi\ldots(P\chi P)\ldots)}_{n-1 \text{ fois}}).$$

De même la probabilité $Q^{(n)}$ est décrite par

$$L_0 \chi(\underbrace{Q\chi(Q\chi\ldots\chi(Q\chi Q)\ldots)}_{n-1 \text{ fois}})$$

On en déduit que la mesure $P^{(n)^{1/2}} Q^{(n)^{1/2}}$ est décrite par

$$L_0 \chi(\underbrace{\mathcal{R}\chi(\mathcal{R}\chi\ldots\chi(\mathcal{R}\chi\mathcal{R})\ldots)}_{n-1 \text{ fois}}).$$

Donc $\rho(P^{(n)}, Q^{(n)}) = \Sigma\{\sigma[L_0 \chi(\mathcal{R}\chi(\mathcal{R}\chi\ldots\chi(\mathcal{R}\chi\mathcal{R})\ldots))]\} = \Sigma(\sigma(L_0 \chi \mathcal{R}^n))$.

Donc d'après le lemme 5 de l'appendice, $\rho(P^{(n)}, Q^{(n)}) \underset{n \to +\infty}{\to} 0 \Longleftrightarrow \mathcal{R}^n \underset{n \to +\infty}{\to} 0$

On en déduit facilement que s'il existe un vecteur colonne L tel que $\mathcal{R}L = L$ alors $L = 0$ ce qui entraine encore que $\det (I - \mathcal{R}) \neq 0$ et réciproquement.

Nous pouvons alors énoncer le résultat suivant :

18 - Théorème : Soient (X_n) et (Y_n) deux suites de variables aléatoires définies sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans $(\mathcal{X}, \mathfrak{B})$ constituant chacune une chaîne de Markov homogène de matrices de transition respectives P et Q : une condition nécessaire et suffisante pour qu'il n'y ait pas séparation asymptotique des deux chaînes est qu'il existe un sous-ensemble clos commun à P et Q tel que la restriction de P à ce sous-ensemble soit égale à la restriction de Q à ce sous-ensemble.

Supposons tout d'abord qu'il existe un sous-ensemble clos commun à P et Q sur lequel P et Q coïncident. Les matrices P et Q s'écrivent

$$P = \begin{pmatrix} P_{\alpha\alpha} & 0 \\ P_{\beta\alpha} & P_{\beta\beta} \end{pmatrix} \qquad Q = \begin{pmatrix} P_{\alpha\alpha} & 0 \\ Q_{\beta\alpha} & Q_{\beta\beta} \end{pmatrix}$$

avec $\alpha+\beta = m$.

On en déduit que
$$\mathcal{R} = \begin{pmatrix} P_{\alpha\alpha} & 0 \\ * & * \end{pmatrix}$$

où l'astérisque représente une sous-matrice quelconque. Deux astérisques figurant dans une partition de matrice ou dans des partitions de deux matrices différentes ne représentent pas nécessairement la même sous-matrice.

Alors
$$\mathcal{R}^n = \begin{pmatrix} P^n_{\alpha\alpha} & 0 \\ * & * \end{pmatrix}$$

D'après le corollaire 9, $||\mathcal{R}^n|| \geq ||P^n_{\alpha\alpha}||$.
La matrice $P_{\alpha\alpha}$ étant stochastique, la matrice $P^n_{\alpha\alpha}$ l'est également, donc $||P^n_{\alpha\alpha}|| = 1$. On en déduit que $||\mathcal{R}^n|| = 1$. Il n'y a donc pas séparation asymptotique.

Inversement supposons qu'il n'y ait pas séparation asymptotique.

Alors d'après le théorème 17 det $(I - \mathcal{R}) = 0$. Donc la matrice sous Markovienne $\mathcal{R}$ qui admet 1 pour valeur propre contient d'après la proposition 15 une sous-matrice stochastique.

$$\mathcal{R} = \begin{pmatrix} \mathcal{R}_{\alpha\alpha} & 0 \\ * & * \end{pmatrix} \quad \text{où } \mathcal{R}_{\alpha\alpha} \text{ est stochastique}$$

On en déduit à l'aide du lemme (14) que

$$P = \begin{pmatrix} \mathcal{R}_{\alpha\alpha} & 0 \\ * & * \end{pmatrix} \quad \text{et} \quad Q = \begin{pmatrix} \mathcal{R}_{\alpha\alpha} & 0 \\ * & * \end{pmatrix}$$

d'où le résultat.

19 - <u>Corollaire</u> : <u>en reprenant les hypothèses du théorème précédent, une condition nécessaire et suffisante pour qu'il n'y ait pas séparation asymptotique est que</u> P <u>et</u> Q <u>admettent une classe finale commune sur laquelle elles coïncident.</u>

Preuve : Tout d'abord une classe finale est un ensemble clos.

D'autre part tout sous-ensemble clos contient une classe finale.

20 - <u>Corollaire</u> : <u>en reprenant les hypothèses du théorème</u> 18

<u>Si les matrices</u> P <u>et</u> Q <u>sont irréductibles et si</u> $P \neq Q$ <u>il y a séparation asymptotique</u>.

21 - <u>Exemples</u> : Considérons les deux matrices stochastiques irréductibles suivantes :

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}$$

$$Q = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & a & 1-a & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}$$

On vérifie facilement que $\det(I - P) = \det(I - Q) = 0$ où I est la matrice identité de $\mathbb{R}^5$.

On a d'autre part

$$\mathcal{R} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & \frac{a}{2} & \frac{1-a}{2} & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}$$

$\det(I - \mathcal{R}) = \frac{1}{16}\left(1 - \sqrt{\frac{a}{2}} - \sqrt{\frac{1-a}{2}}\right)$. On en déduit que :

$$\det(I - \mathcal{R}) = 0 \Longleftrightarrow 1 = \sqrt{\frac{a}{2}} + \sqrt{\frac{1-a}{2}} \Longleftrightarrow \sqrt{a(1-a)} = \frac{1}{2} \Longleftrightarrow a = \frac{1}{2}$$

22 - On étudie ici d'une autre façon les cas $m = 2$ et $m = 3$, ce qui nous permet de retrouver le théorème 18 dans ces cas particuliers.

<u>Cas m = 2.</u> Considérons les matrices P et Q données par :

$$P = \begin{pmatrix} a_1 & 1-a_1 \\ a_2 & 1-a_2 \end{pmatrix} \qquad Q = \begin{pmatrix} b_1 & 1-b_1 \\ b_2 & 1-b_2 \end{pmatrix}$$

Supposons qu'il n'y ait pas séparation asymptotique.

D'après le lemme 10 et le résultat 17 il faut que $||\mathcal{R}|| = 1$

Ceci impose d'après le lemme 14 que P et Q coïncident sur au moins une ligne.

Supposons que ce soit la première : $a_1 = b_1$.

Si $a_1 = 1$ le théorème 18 est obtenu puisque $\{1\}$ est un état absorbant commun à P et Q.

Sinon on a :

$$\mathcal{R} = \begin{pmatrix} a_1 & 1-a_1 \\ \sqrt{a_2 b_2} & \sqrt{(1-a_2)(1-b_2)} \end{pmatrix}$$

$$I - \mathcal{R} = \begin{pmatrix} 1-a_1 & -(1-a_1) \\ -\sqrt{a_2 b_2} & 1-\sqrt{(1-a_2)(1-b_2)} \end{pmatrix}$$

$$\det(I - \mathcal{R}) = (1-a_1)\left[(1 - \sqrt{(1-a_2)(1-b_2)} - \sqrt{a_2 b_2}\right] = 0$$

$$\Leftrightarrow \quad 1 - \sqrt{(1-a_2)(1-b_2)} = \sqrt{a_2 b_2}$$

Elevons au carré. $1 + (1-a_2)(1-b_2) - 2\sqrt{(1-a_2)(1-b_2)} = a_2 b_2$

$$2 + a_2 b_2 - a_2 - b_2 - 2\sqrt{(1-a_2)(1-b_2)} = a_2 b_2$$

$$\Leftrightarrow \quad 2\sqrt{(1-a_2)(1-b_2)} = 2 - a_2 - b_2$$

élevons au carré, il vient $(a_2 - b_2)^2 = 0$

$\Leftrightarrow$ $a_2 = b_2$ donc $P = Q$ et le théorème est vérifié.

<u>Cas m = 3.</u> Supposons qu'il n'y ait pas séparation asymptotique : les mêmes considérations que pour m = 2 nous indiquent qu'il existe au moins une ligne sur laquelle P et Q coïncident.

Supposons qu'il s'agisse de la première.

Si $a_{11} = b_{11} = 1$ le théorème est démontré, car $\{1\}$ constitue un état absorbant commun à P et Q. Sinon : $a_{11} \neq 1$.

$$P = \begin{pmatrix} a_{11} & a_{12} & 1-(a_{11}+a_{12}) = a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$Q = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

$$\mathcal{R} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ \sqrt{b_{21}a_{21}} & \sqrt{b_{22}a_{22}} & \sqrt{a_{23}b_{23}} \\ \sqrt{a_{31}b_{31}} & \sqrt{a_{32}b_{32}} & \sqrt{a_{33}b_{33}} \end{pmatrix}$$

Supposons qu'il n'ait pas d'autres lignes que la première sur lesquelles P et Q coïncident.

Alors en calculant $\mathcal{R}^2$ on voit que $||\mathcal{R}^2|| < 1$ et donc $\lim\limits_{n \to +\infty} ||\mathcal{R}^n|| = 0$ ce qui est contraire à l'hypothèse.

Il existe donc au moins deux lignes sur lesquelles P et Q coïncident.

On peut alors écrire $\mathcal{R}$ de la façon suivante :

$$\mathcal{R} = \begin{pmatrix} 1-(a_{12}+a_{13}) & a_{12} & a_{13} \\ a_{21} & 1-(a_{21}+a_{23}) & a_{23} \\ \sqrt{a_{31}b_{31}} & \sqrt{a_{32}b_{32}} & \sqrt{a_{33}b_{33}} \end{pmatrix}$$

$$I-\mathcal{R} = \begin{pmatrix} a_{12}+a_{13} & -a_{12} & -a_{13} \\ -a_{21} & a_{21}+a_{23} & -a_{23} \\ -\sqrt{a_{31}b_{31}} & -\sqrt{a_{32}b_{32}} & 1-\sqrt{a_{33}b_{33}} \end{pmatrix}$$

On calcule det $(I-\mathcal{R})$ en ajoutant la dernière colonne à la seconde puis la première colonne à la seconde. On développe alors suivant la deuxième colonne

et l'on trouve :

$$\det(I - R) = 0 \iff$$

$$(a_{12}a_{23} + a_{13}a_{23} + a_{21}a_{13})(1-\sqrt{a_{31}b_{31}} - \sqrt{a_{32}b_{32}} - \sqrt{a_{33}b_{33}}) = 0$$

Dans ce cas

ou $$(a_{12}a_{23} + a_{13}a_{23} + a_{21}a_{13}) = 0$$

ce qui veut dire que :

$$\underline{\text{soit}}\ a_{13} = 0 \quad \text{et} \begin{cases} a_{12} = 0 \\ \text{ou} \\ a_{23} = 0 \end{cases}$$

$a_{12} = 0$ est contraire à l'hypothèse car alors $a_{11} = 1$ donc $a_{13} = a_{23} = 0$ et

$$P = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ & * & \end{pmatrix} \quad \text{et } Q = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ & * & \end{pmatrix}$$

et le résultat est démontré ;

$$\underline{\text{soit}}\ a_{23} = 0 \quad \text{et} \begin{cases} a_{21} = 0 \\ \text{ou} \\ a_{13} = 0 \end{cases}$$

Si $a_{23} = 0 = a_{21}$ alors $a_{22} = 1$ et le résultat est démontré

ou $a_{23} = a_{13} = 0$ et P est encore de la forme ci-dessus

ou $$(1 - \sqrt{a_{31}b_{31}} - \sqrt{a_{32}b_{32}} - \sqrt{a_{33}b_{33}}) = 0 \iff \sum_{j=1}^{3} \sqrt{a_{3j}b_{3j}} = 1$$

On sait grâce à l'inégalité de Cauchy-Schwarz que :

$$\sum_{j=1}^{3} \sqrt{a_{3j}b_{3j}} \leq (\sum_{j=1}^{3} a_{3j})^{1/2} (\sum_{j=1}^{3} b_{3j})^{1/2} = 1$$

D'autre part l'égalité est vérifiée si et seulement si : $\exists\, \lambda \in \mathbb{R}$ tel que

$$\begin{pmatrix} a_{31} \\ a_{32} \\ a_{33} \end{pmatrix} = \lambda \begin{pmatrix} b_{31} \\ b_{32} \\ b_{33} \end{pmatrix}$$

ce qui conduit à $\lambda = 1$ et à $\forall\ j = 1,\ldots,3$ $a_{3j} = b_{3j}$.

Alors $P = Q$. On en conclut le théorème.

§ III - 4 Problèmes liés à l'estimation du paramètre θ dans les chaînes de Markov et la séparation asymptotique

On suppose que pour tout $\theta \in \Theta$ sous-ensemble ouvert de $\mathbb{R}^r$ (ou espace euclidien de dimension r) $(p_{ij}(\theta))$ est une matrice stochastique mxm. Lorsqu'on ne fait aucune hypothèse sur la régularité des fonctions p_{ij}, on a vu (cf. chap. III-3) les conditions nécessaires et suffisantes de séparation asymptotique ou non séparation asymptotique point par point.

On va donc faire des hypothèses (très classiques) de régularité des fonctions p_{ij} et en mesurer les conséquences en termes d'estimation et de séparation asymptotique.

On suppose donc que Θ est un ouvert de $\mathbb{R}^r$ muni de la mesure Lebesgue et que pour tout i, j:$p_{ij}(\theta)$ admet des dérivées partielles du premier ordre. On notera D la matrice m^2xr dont les éléments sont

$$d_{iju} = \frac{\partial p_{ij}(\theta)}{\partial \theta_u} \qquad 1 \leq u \leq r$$

On supposera $r \leq m^2-m$

On veut établir les liens entre la séparation asymptotique et l'expression de la matrice D.

Notons que le rang de la matrice D est lié plus directement à une notion de dépendance des (p_{ij}) : i.e. si D est de rang < r l'une au moins des p_{ij} se déduit des autres. Mais la notion de séparation est liée a une notion de fixité des fonctions p_{ij} sur une classe finale.

Faisons l'hypothèse de Billingsley $p_{ij}(\theta) > 0$ pour tout i et j.

23 - Lemme : s'il existe un sous-ensemble de mesure de Lebesgue non nulle Θ_0 de Θ tel que pour tout θ^0, $\theta^1 \in \Theta_0$ il n'y a pas de séparation asymptotique des chaînes de Markov associées, alors la matrice D est nulle sur un sous-ensemble de mesure non nulle.

Preuve : on peut donc trouver $\theta^0 \in \Theta_0$ et une boule ouverte centrée en θ^0 et

incluse dans Θ_0 tels que $\forall\ \theta \in B(\theta^0,\varepsilon_0)$ les matrices $(p_{ij}(\theta^0))$ et $(p_{ij}(\theta))$ ne se séparent pas asymptotiquement.

Donc puisque les matrices en question sont irréductibles $\forall i, j, \forall\ \theta \in B(\theta^0,\varepsilon_0)$

$$p_{ij}(\theta) = p_{ij}(\theta^0)$$

donc sur $B(\theta^0,\varepsilon_0)$

$$\frac{\partial p_{ij}}{\partial \theta_u} = 0 \text{ pour } u = 1,\dots,r.$$

24 - <u>Corollaire</u> : <u>Si pour tout</u> $\theta \in \Theta$ <u>la matrice</u> D <u>est de rang</u> r, <u>alors mis à part sur un sous-ensemble de mesure nulle, il y a séparation asymptotique des chaînes de Markov de matrices</u> $(p_{ij}(\theta))$, $(p_{ij}(\theta'))$ $\theta, \theta \in \Theta$.

Remarque : on voit donc que Billingsley [2] faisait dans son théorème 4-1. des hypothèses qui garantissaient non seulement l'indépendance des p_{ij} mais aussi la séparation asymptotique presque partout. En supposant de plus que les p_{ij} sont de classe $\mathcal{C}^2$ ou $\mathcal{C}^3$ il construit une suite $\hat{\theta}$ qui converge en probabilité vers θ.

Pour faire une étude dans le cas général, il faut supposer $p_{ij}(\theta) \geq 0$. La nouveauté est donc d'autoriser les fonctions p_{ij} à être nulles, hypothèse rarement faite dans la théorie classique de l'estimation (cf par ex. Cramer [5], Billingsley [1] ou [2]).Pour qu'il y ait séparation asymptotique, on a vu qu'il suffit qu'il existe au moins un élément de chanque classe finale qui diffère d' une chaîne à l'autre.

1°) Si $\forall\ i = 1,\dots,m$ les fonctions p_{ij} sont bijectives : il y aura toujours séparation asymptotique.

2°) Si $\{p_{ij}/p_{ij} = 0\}$ ne dépend pas de θ, alors le nombre de classes finales du processus est déterminé une fois pour toute. On fera les hypothèses de régularité au niveau de chaque classe finale.

3°) cas général : l'élargissement de l'hypothèse sur les p_{ij} à la nullité éventuelle va nous conduire à resserver quelque peu l'hypothèse sur le rang de la matrice D.

On propose le résultat suivant :

25 - Théorème : Soit Θ un sous-ensemble borné de $\mathbb{R}^r$.

Soit $I_k = \{i_1,\ldots,i_k\}$ un sous ensemble quelconque de $\{1,\ldots,m\}$. Si pour tout $i_k \in I_k$ $\exists$ j_k $\{1,\ldots,m\}$ dépendant éventuellement de I_k tel que

$$\frac{D(p_{i_h j_k} \,/\, h = 1\ldots k)}{D(\theta_1,\ldots,\theta_r)}$$ est pour tout θ de rang $\min(k, r)$, et

ceci pour tout k alors pour tout $\theta \in \Theta$, l'ensemble des $\theta' \in \Theta$ pour lesquels il n'y a pas de séparation asymptotique est de mesure nulle.

Démonstration : plaçons nous au voisinage d'un point θ_0 pour lequel P_{θ_0} n'est pas irréductible. Eventuellement à la suite d'une nouvelle numérotation de l'espace des états, on peut écrire P_{θ_0} sous la forme :

$$P_{\theta_0} = \begin{pmatrix} P_{kk} & 0 \\ P_{m-k,k} & P_{m-k,m-k} \end{pmatrix}$$

où la matrice P_{kk} est irréductible.

Posons donc $p_{ij}(\theta) = \alpha_{ij}$ si $\begin{cases} 1 \le i \le k \\ 1 \le j \le k \end{cases}$

on a d'autre part $p_{ij}(\theta) = 0$ $\begin{cases} 1 \le i \le k \\ k+1 \le j \le m \end{cases}$

Considérons le système (26)

$$(26)\qquad \left.\begin{array}{ll} p_{ij} - \alpha_{ij} = 0 & \\ 1 \le j \le k & \\ p_{ij} \qquad\quad = 0 & \\ k+1 \le j \le m & \end{array}\right\} \qquad 1 \le i \le k$$

L'ensemble des θ qui vérifient le système (1) est fermé comme intersection finie non vide des fermés puisque les p_{ij} sont continues et que θ_0 au moins vérifie ce système.

Cet ensemble est donc compact puisque Θ est borné.

Par hypothèse $\exists\ j_1,\ldots,j_k$ tels que

$$\frac{D\ (p_{kj_h}\ /h = 1\ldots k)}{D\ (\theta_1,\ldots,\theta_r)}$$

est de rang min (k, r) pour tout θ.

On peut donc trouver $s < r$ tel que

$$\frac{D\ (p_{kj_h}\ /\ h = 1\ldots s)}{D(\theta_{i_1},\ldots,\theta_{i_s}\ )} \neq 0$$

$s = k$ si $k < r$ sinon par hypothèse $k = r$ et l'on peut trouver un mineur non nul du déterminant $\frac{D(p)}{D(\theta)}$ d'ordre $s < r$

Le système (26) peut donc s'écrire

$$(26')\qquad \left.\begin{array}{ll} p_{ij}(a,\ b) - \alpha_{ij} = 0 & \\ 1 \le j \le k & \\ p_{ij}(a,\ b) \qquad\quad = 0 & \\ k+1 \le j \le m & \end{array}\right\} \qquad 1 \le i \le k$$

où $\qquad b = \theta_{i_1},\ldots,\theta_{i_s}$

$$a = \theta_{j_1},\ldots,\theta_{j_{r-s}} \quad \text{où} \quad j_h \neq i_l \quad \forall\, l = 1\ldots s \quad \forall\, h = 1,\ldots,r-s$$

par application du théorème des fonctions implicites, il existe un voisinage A de a et B de b tel que le système (26') détermine implicitement une application unique φ de A dans B avec $\varphi(a) = b$.

On peut donc faire un recouvrement de l'ensemble des θ qui vérifient le système (26) par des ouverts du type A x B et en extraire un sous recouvrement fini.

On aura donc un nombre fini de fonction φ_t vérifiant une équation du type $\varphi_t(a) = b$; chacune caractérisant une sous-variété de $\mathbb{R}^r$ de dim < r-s < r donc de mesure de Lebesgue sur $\mathbb{R}^r$ nulle. Leur union finie est donc de mesure nulle. Pour tout θ l'ensemble des θ' tels que les chaînes associées à P_θ et $P_{\theta'}$ ne se séparent pas asymptotiquement est donc de mesure nulle.

Si l'on est au voisinage d'un point θ pour lequel la matrice P_θ est irréductible on applique la démonstration précédente à P_θ.

27 - <u>Autre approche du problème</u> : considérons le cas m = 2 = k et supposons que l'on définisse les fonctions suivantes :

$$\underline{\theta} = (\theta_1, \theta_2) \longrightarrow z = \theta_1 + i\theta_2 \in \mathbb{C} \xrightarrow{f_{ij}} f_{ij}(z) \in \mathbb{C}$$

$$\longrightarrow p_{ij}(\underline{\theta}) = |f_{ij}(z)|$$ où Θ est un ouvert précompact de $\mathbb{R}^2$ et $f_{ij}(\Theta) \subset \{ z \in \ /|z| \leq 1 \}$, enfin $\underline{\theta} \neq \underline{\theta}' \Rightarrow P_{\underline{\theta}} \neq P_{\underline{\theta}'}$

Comme $|f_{12}| = 1 - |f_{11}|$ on ne parlera que de f_1 et f_2

$$P_{\underline{\theta}} = \begin{pmatrix} |f_1(\underline{\theta})| & 1 - |f_1(\underline{\theta})| \\ 1 - |f_2(\theta)| & |f_2(\theta)| \end{pmatrix}$$

On suppose f_i $(i = 1,2)$ définie et continue sur $D = \{z \,/\, z = \theta_1 + i\theta_2$ où $(\theta_1, \theta_2) = \underline{\theta}\}$ et que f_i possède la propriété de la moyenne sur D.

On supposera de plus que D est un ouvert connexe du plan $\mathbb{C}$. (D est borné puisque Θ est précompact). Soit $\overline{D}$ l'adhérence de D. Soit M la borne supérieure de $|f_i(z)|$ quand z parcourt la frontière de D. On aura $M \leq 1$ par hypothèse. Comme corollaire du principe du maximum (cf [3]) $\forall\, z \in D$, $|f_i(z)| \leq 1$. Et s'il existe $z_0 \in D$ tel que $|f_i\ (z_0)| = 1$ alors f_i est constante sur D : il n'y a jamais séparation asymptotique dans ce cas.

On supposera donc $|f_i(z)| < 1$ sur D. Donc la matrice $P_{\underline{\theta}}$ est irréductible $\forall\, \underline{\theta} \in \Theta$. A l'aide de la condition $\underline{\theta} \neq \underline{\theta}' \Longrightarrow P_{\underline{\theta}} \neq P_{\underline{\theta}'}$, on en déduit qu'il y a toujours séparation asymptotique.

APPENDICE

1 - <u>Définition</u> : on appelle matrice carrée d'ordre m à sous-élements d'ordre m^k une matrice A mxm dont les élements a_{ij} sont des vecteurs $\in \mathbb{R}^{m^k}$. les composantes de ces vecteurs sont appelés des sous-élements.

On voit qu'on peut identifier une telle matrice à un élément de $\mathbb{R}^{m^{k+2}}$. On dira que A est une matrice (m^k,m)xm. On peut identifier une matrice carrée ordinaire à une matrice (1,m)xm. On peut dédinir sur les matrices à sous-éléments les opérations habituelles de somme, produit scalaire.

2 - <u>Définition du produit matriciel</u> .

Soient A et B deux matrices à sous-éléments : A : (m^k,m)xm et B : (m^l,m)xm. On définit sur ces matrice le produit $C = A \chi B$ par

$$C = (C_{ij}) \quad \begin{matrix} i = 1\ldots m \\ j = 1\ldots m \end{matrix}$$

où $C_{ij} = a_{i1} \otimes b_{1j}, \ldots, a_{im} \otimes b_{mj}$

avec $a_{ik} \otimes b_{kj} = a^1_{ik}\, b_{1j}, \ldots, a^{m^k}_{ik}\, b_{kj}$

La matrice C est alors une matrice à sous-éléments (m^{k+l+1},m)xm. Le produit noté χ est appelé produit décomposé de matrices à sous-éléments.

N.B. il ne s'agit pas, a un isomorphisme près, du produit tensoriel de deux matrices, comme on peut le voir de la façon suivante : le produit tensoriel d'une matrice 2x2 par une matrice 2x2 est une matrice 4x4, alors que le produit décomposé d'une matrice 2x2 par une matrice 2x2 est une matrice 4x2 à un isomorphisme près.

On vérifie que ce produit n'est pas associatif, car bien que les sous-éléments du produit $(A \chi B) \chi D$ soient les mêmes que ceux du produit $A \chi (B \chi D)$ l'ordre des sous-éléments est permuté au sein d'un même élément.

Pour une matrice carrée ordinaire mxm : $B = (b_{ij})$, on note

$$\Sigma B = \sum_{\substack{i=1\ldots m \\ j=1\ldots m}} b_{ij}$$

Pour une matrice à sous-éléments A (m^k,m)xm $A = (a_{ij})$ $a_{ij} \in \mathbb{R}^{m^k}$ on définit : l'opérateur σ : σA est la matrice obtenue en faisant la somme des sous-éléments d'un même élément. On obtient donc une matrice carrée mxm.

Si A et B sont deux matrices carrées mxm ordinaires on a immédiatement :

$$\sigma(A \chi B) = A\ B$$

où A B est le produit matriciel ordinaire.

Si A et B sont deux matrices à sous-éléments (m^k,m)xm et (m^l,m)xm on a :

$$\sigma(A \chi B) = \sigma(\sigma A \chi B) = \sigma(A \chi \sigma B)$$

On en déduit que $\sigma(\underbrace{P \chi (P \chi \ldots \chi P)\ldots}_{n \text{ fois}})) = \sigma(P \chi \sigma P^{n-1}) = P^n$

Comme on récupère l'associativité via l'opération σ on pourra parler de $\underbrace{\sigma(P \chi \ldots \chi P)}_{n \text{ fois}}$

3 - <u>Proposition</u>

<u>Soit</u> P <u>la matrice carrée mxm de transition d'une chaîne de Markov homogène</u>. <u>Si l'on fait</u>

$$(\underbrace{P \chi (P \chi \ldots (P \chi P)\ldots}_{n \text{ fois}}))$$

<u>on décrit la loi de</u> $(X_l, \ldots X_{l+n} \ / \ X_{l-1})$.

<u>L'élément qui se trouve à l'intersection de la ligne</u> i <u>et la colonne</u> j <u>donne les probabilités de tous les états</u> $X_l \ldots X_{l+n}$ <u>sachant</u> $X_{l-1} = i$ <u>et tels que</u> $X_{l+n} = j$.

Se démontre par récurrence.

4 - Corollaire

Appelons L_0 le vecteur colonne à m composantes qui décrit la loi initiale de la chaîne de Markov. (loi de X_1). Alors

$$L_0' \chi(\underbrace{P\chi(P\chi\ldots(P\chi P)\ldots))}_{n-1 \text{ fois}}$$

décrit la loi conjointe du vecteur $(X_1,\ldots, X_n)$.

On trouve dans l'ordre pour $i = 1,\ldots m$ les probabilités conjointes de X_1 en i et terminant par X_n en 1, puis les probabilités de X_1 en i pour $i = 1,\ldots m$ et terminant par X_n en 2 ; ainsi de suite jusqu'à m.

5 - Lemme : Soit L un vecteur de $\mathbb{R}^m$ fixé de composantes > 0.

Alors $\quad \Sigma(\sigma(L'\chi A^n) \underset{n\to+\infty}{\longrightarrow} 0 \iff A^n \underset{n\to+\infty}{\longrightarrow} 0$

Preuve : $\quad \Sigma(\sigma(L'\chi A\,\chi(A\ldots\chi(A\chi A)\ldots))) = \Sigma(\sigma(L'\chi A^n)) \underset{n\to+\infty}{\longrightarrow} 0 \iff$

$\sigma(L'\chi A^n) \underset{n\to+\infty}{\longrightarrow} 0 \Leftrightarrow A^n \to 0$

BIBLIOGRAPHIE DU CHAPITRE III

[1] Billingsley : Statistical inference for Markov processes.

[2] Billingsley : Statistical methods in Markov chains.

[3] Cartan : Théorie des fonctions d'une ou plusieurs variables complexes.

[4] K.L. CHUNG : Markov chains with stationary transition probabilities. Springer V. 1960

[5] Cramer : Mathematical methods of statistics.

[6] A.R. Darwich : Propriétés de séparation asymptotique dans les chaînes de Markov. Thèse - Paris 1982.

[7] J.G. Kemeny et J.L. Snell : finite Markov chains. Springer 1976

[8] S. Karlin : Initiation aux processus aléatoires. Dunod 1969

[9] A. Hillion : Quelques questions de distances de lois de probabilités. Publications I.S.U.P. Vol. XXIII, 1978, fasc.3-4 pp 37-38

[10] A. Hillion : Etude du point de vue de la séparation asymptotique ... d'estimation ponctuelle. Thèses Paris 1980.

[11] A. Kolmogorov - S. Fomine : Eléments de la théorie des fonctions et de l'analyse fonctionnelle. (Traduction française - M.I.R. 1974)

[12] R. Moché : "Décantation et séparation asymptotique uniforme ; test et estimateurs convergents dans le cas d'observations indépendantes, équidistribuées ou non". Thèse - Lille (1977).

[13] R. Pallu de la Barriere : Cours d'automatique théorique - Dunod (1966).

dernière page de la thèse

AUTORISATION DE SOUTENANCE

VU les dispositions de l'article 5 de l'arrêté du 16 Avril 1974,

VU les rapports de M. Jean GEFFROY

M. Denis BOSQ

M. Bernard VAN CUTSEM

M. Bernard GAREL est autorisé à présenter une thèse en soutenance pour l'obtention du grade de DOCTEUR D'ETAT ES SCIENCES.

Fait à GRENOBLE, le

Le Président de l'U.S.M.G.

Vice-Président
M. TANCHE

Le Président de l'I.N.P.G.

D. BLOCH
Président
de l'Institut National Polytechnique
de Grenoble

P.O. le Vice-Président,