



**HAL**  
open science

# Détection des valeurs aberrantes dans un échantillon gaussien multidimensionnel

Bernard Garel

► **To cite this version:**

Bernard Garel. Détection des valeurs aberrantes dans un échantillon gaussien multidimensionnel. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG; Université Joseph-Fourier - Grenoble I, 1976. Français. NNT: . tel-00287197

**HAL Id: tel-00287197**

**<https://theses.hal.science/tel-00287197>**

Submitted on 11 Jun 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

**UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE**  
**INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE**

POUR OBTENIR LE GRADE DE  
Docteur de 3ème cycle  
en MATHÉMATIQUES APPLIQUÉES

**Bernard GAREL**

**DETECTION DES VALEURS ABERRANTES  
DANS UN ECHANTILLON GAUSSIEN  
MULTIDIMENSIONNEL**

soutenue le 28 juin 1976 devant la commission d'examen

Président: J.R. BARRA

J. GEFFROY

Examineurs: G. ROMIER

B. VAN CUTSEM

**THESE**  
PREPAREE A LA  
FACULTE DES SCIENCES ET DES TECHNIQUES DE CHAMBERY



UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

Monsieur Michel SOUTIF : Président

Monsieur Gabriel CAU : Vice-Président

---

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

PROFESSEURS TITULAIRES

MM.	ANGLES D'AURIAC Paul	Mécanique des Fluides
	ARNAUD Paul	Chimie
	AUBERT Guy	Physique
	AYANT Yves	Physique Approfondie
Mme	BARBIER Marie-Jeanne	Electrochimie
MM.	BARBIER Jean-Claude	Physique Expérimentale
	BARBIER Reynold	Géologie Appliquée
	BARJON Robert	Physique Nucléaire
	BARNOUD Fernand	Biosynthèse de la Cellulose
	BARRA Jean-René	Statistiques
	BARRIE Joseph	Clinique Chirurgicale
	BEAUDOING André	Clinique de Pédiatrie et Puériculture
	BERNARD Alain	Mathématiques Pures
Mme	BERTRANDIAS Françoise	Mathématiques Pures
MM.	BERTRANDIAS Jean-Paul	Mathématiques Pures
	BEZES Henri	Pathologie Chirurgicale
	BLAMBERT Maurice	Mathématiques Pures
	BOLLIER Louis	Informatique (I.U.T. B)
	BONNET Georges	Electrotechnique
	BONNET Jean-Louis	Clinique Ophtalmologique
	BONNET-EYMARD Joseph	Clinique Gastro-entérologique
Mme	BONNIER Marie-Jeanne	Chimie Générale
MM.	BOUCHERLE André	Chimie et Toxicologie
	BOUCHEZ Robert	Physique nucléaire
	BOUSSARD Jean-Claude	Mathématiques Appliquées
	BRAVARD Yves	Géographie
	CABANEL Guy	Clinique Rhumatologique et Hydrologique
	CALAS François	Anatomie
	CARLIER Georges	Biologie Végétale
	CARRAZ Gilbert	Biologie Animale et Pharmacodynamie
	CAU Gabriel	Médecine Légale et Toxicologie
	CAUQUIS Georges	Chimie Organique
	CHABAUTY Claude	Mathématiques Pures
	CHARACHON Robert	Clinique Oto-Rhino-Laryngologique
	CHATEAU Robert	Clinique de Neurologie
	CHIBON Pierre	Biologie Animale
	COEUR André	Pharmacie Chimique et Chimie Analytique
	CONTAMIN Robert	Clinique Gynécologique
	COUDERC Pierre	Anatomie Pathologique
	CRAYA Antoine	Mécanique
Mme	DEBELMAS Anne-Marie	Matière Médicale
MM.	DEBELMAS Jacques	Géologie Générale
	DEGRANGE Charles	Zoologie
	DELORMAS Pierre	Pneumo-Physiologie
	DEPORTES Charles	Chimie Minérale
	DESRE Pierre	Métallurgie

MM.	DESSAUX Georges	Physiologie Animale
	DODU Jacques	Mécanique Appliquée (I.U.T. A)
	DOLIQUE Jean-Michel	Physique des Plasmas
	DREYFUS Bernard	Thermodynamique
	DUCROS Pierre	Cristallographie
	DUGOIS Pierre	Clinique de Dermatologie et Syphiligraphie
	GAGNAIRE Didier	Chimie Physique
	GALLISSOT François	Mathématiques Pures
	GALVANI Octave	Mathématiques Pures
	GASTINEL Noël	Analyse Numérique
	GAVEND Michel	Pharmacologie
	GEINDRE Michel	Electroradiologie
	GERBER Robert	Mathématiques Pures
	GERMAIN Jean-Pierre	Mécanique
	GIRAUD Pierre	Géologie
	JANIN Bernard	Géographie
	KAHANE André	Physique Générale
	KLEIN Joseph	Mathématiques Pures
	KOSZUL Jean-Louis	Mathématiques Pures
	KRAVTCHEKNO Julien	Mécanique
	KUNTZMANN Jean	Mathématiques Appliquées
	LACAZE Albert	Thermodynamique
	LACHARME Jean	Biologie Végétale
Mme	LAJZEROWICZ Janine	Physique
MM.	LAJZEROWICZ Joseph	Physique
	LATREILLE René	Chirurgie Générale
	LATURAZE Jean	Biochimie Pharmaceutique
	LAURENT Pierre-Jean	Mathématiques Appliquées
	LEDRU Jean	Clinique Médicale B
	LLIBOUTRY Louis	Géophysique
	LOISEAUX Pierre	Sciences Nucléaires
	LONGEQUEUE Jean-Pierre	Physique Nucléaires
	LOUP Jean	Géographie
Melle	LUTZ Elisabeth	Mathématiques Pures
	MALGRANGE Bernard	Mathématiques Pures
	BOUTET DE MONVEL Louis	Mathématiques Pures
	MALINAS Yves	Clinique Obstétricale
	MARTIN-NOEL Pierre	Sémiologie médicale
	MAZARE Yves	Clinique Médicale A
	MICHEL Robert	Minéralogie et Pétrographie
	MICOUD Max	Clinique Maladies Infectieuses
	MOURIQUAND Claude	Histologie
	MOUSSA André	Chimie Nucléaire
	MULLER Jean-Michel	Thérapeutique (Néphrologie)
	NEEL Louis	Physique du solide
	OZENDA Paul	Botanique
	PAYAN Jean-Jacques	Mathématiques Pures
	PEBAY-PEYROULA Jean-Claude	Physique
	RASSAT André	Chimie Systématique
	RENARD Michel	Thermodynamique
	RINALDI Renaud	Physique
	DE ROUGEMONT Jacques	Neuro-Chirurgie
	SEIGNEURIN Raymond	Microbiologie et Hygiène
	SENGEL Philippe	Zoologie
	SIBILLE Robert	Construction Mécanique (I.U.T. A)
	SOUTIF Michel	Physique Générale
	TANCHE Maurice	Physiologie

MM.	TRAYNARD Philippe	Chimie Générale
	VAILLANT François	Zoologie
	VALENTIN Jacques	Physique Nucléaire
	VAUQUOIS Bernard	Calcul Electronique
Mme	VERAIN Alice	Pharmacie Galénique
MM.	VERAIN André	Physique
	VEYRET Paul	Géographie
	VIGNAIS Pierre	Biochimie médicale
	YOCCOZ Jean	Physique Nucléaire Théorique

#### PROFESSEURS ASSOCIES

MM.	CLARK Gilbert	Spectrométrie Physique
	CRABBE Pierre	CERMO
	ENGLMAN Robert	Spectrométrie Physique
	HOLTZBERG Frédéric	Basses Températures
	ROST Ernest	Sciences Nucléaires

#### PROFESSEURS SANS CHAIRE

Melle	AGNIUS-DELORD Claudine	Physique Pharmaceutique
	ALARY Josette	Chimie Analytique
MM.	AMBROISE-THOMAS Pierre	Parasitologie
	BELORIZKY Elie	Physique
	BENZAKEN Claude	Mathématiques Appliquées
	BIAREZ Jean-Pierre	Mécanique
	BILLET Jean	Géographie
	BOUCHET Yves	Anatomie
	BRUGEL Lucien	Energétique (I.U.T. A)
	BUISSON René	Physique (I.U.T. A)
	CONTE René	Physique (I.U.T. A)
	DEPASSEL Roger	Mécanique des Fluides
	GAUTHIER Yves	Sciences Biologiques
	GAUTRON René	Chimie
	GIDON Paul	Géologie et Minéralogie
	GLENAT René	Chimie organique
	GROULADE Joseph	Biochimie Médicale
	HACQUES Gérard	Calcul Numérique
	HOLLARD Daniel	Hématologie
	HUGONOT Robert	Hygiène et Médecine Préventive
	IDELMAN Simon	Physiologie Animale
	JOLY Jean-René	Mathématiques Pures
	JULLIEN Pierre	Mathématiques Appliquées
Mme	KAHANE Josette	Physique
MM.	KUHN Gérard	Physique (I.U.T. A)
	LE ROY Philippe	Mécanique (I.U.T. A)
	LUU DUC Cuong	Chimie Organique
	MAYNARD Roger	Physique du Solide
	PELMONT Jean	Biochimie
	PERRIAUX Jean-Jacques	Géologie et Minéralogie
	PSISTER Jean-Claude	Physique du Solide
Melle	PIERY Yvette	Physiologie Animale
MM.	RAYNAUD Hervé	M.I.A.G.
	REBECQ Jacques	Biologie (CUS)
	REVOL Michel	Urologie
	REYMOND Jean-Charles	Chirurgie Générale
	RICHARD Lucien	Biologie Végétale
Mme	RINAUDO Marguerite	Chimie Macromoléculaire
MM.	ROBERT André	Chimie Papetière

MM.	SARRAZIN Roger	Anatomie et Chirurgie
	SARROT-REYNAUD Jean	Géologie
	SIROT Louis	Chirurgie Générale
Mme	SOUTIF Jeanne	Physique Générale
MM.	STREGLITZ Paul	Anesthésiologie
	VIALON Pierre	Géologie
	VAN CUTSEM Bernard	Mathématiques Appliquées

MAITRES DE CONFERENCES ET MAITRES DE CONFERENCES AGREGES

MM.	AMBLARD Pierre	Dermatologie
	ARMAND Gilbert	Géographie
	ARMAND Yves	Chimie (I.U.T. A)
	BACHELOT Yvan	Endocrinologie
	BARGE Michel	Neuro chirurgie
	BARJOLLE Michel	MIAG
	BEGUIN Claude	Chimie organique
Mme	BERTEL Hélène	Pharmacodynamie
MM.	BOST Michel	Pédiatrie
	BOUCHARLAT Jacques	Psychiatrie adultes
Mme	BOUCHE Liane	Mathématiques (C.U.S.)
MM.	BRODEAU François	Mathématiques (I.U.T. B)
	BUTTEL Jean	Orthopédie
	CHAMBAZ Edmond	Biochimie médicale
	CHAMPETIER Jean	Anatomie et Organogénèse
	CHARDON Michel	Géographie
	CHERADAME Hervé	Chimie Papetière
	CHLAVERINA Jean	Biologie Appliquée (EFP)
	COHEN-ADDAD Jean-Pierre	Spectrométrie Physique
	COLOMB Maurice	Biochimie Médicale
	CONTAMIN Charles	Chirurgie thoracique et cardio-vasculaire
	CORDONNIER Daniel	Néphrologie
	COULOMB Max	Radiologie
	CROUZET Guy	Radiologie
	CYROT Michel	Physique du solide
	DELOBEL Claude	M.I.A.G.
	DENIS Bernard	Cardiologie
	DOUCE Roland	Physiologie Végétale
	DUSSAUD René	Mathématiques (C.U.S.)
Mme	ETERRADOSSI Jacqueline	Physiologie
MM.	FAURE Jacques	Médecine Légale
	FAURE Gilbert	Urologie
	FONTAINE Jean-Marc	Mathématiques Pures
	GAUTIER Robert	Chirurgie Générale
	GENSAC Pierre	Botanique
	GIDON Maurice	Géologie
	GROS Yves	Physique (I.U.T. A)
	GUITTON Jacques	Chimie
	HICTER Pierre	Chimie
	IVANES Marcel	Electricité
	JALBERT Pierre	Histologie
	KOLODIE Lucien	Hématologie
	KRAKOWIAK Sacha	Mathématiques Appliquées
	LE NOC Pierre	Bactériologie-virologie
	LEROY Philippe	I.U.T. A
	MACHE Régis	Physiologie Végétale
	MAGNIN Robert	Hygiène et Médecine Préventive
	MALLION Jean-Michel	Médecine du Travail
	MARECHAL Jean	Mécanique (I.U.T. A)
	MARTIN-BOUYER Michel	Chimie (C.U.S.)

M.	MICHOULIER Jean	Physique (I.U.T. A)	
Mme	MINIER Colette	Physique (I.U.T. A)	
MM.	NEGRE Robert	Mécanique (I.U.T. A)	
	NEMOZ Alain	Thermodynamique	
	NOUGARET Marcel	Automatique (I.U.T.A)	
	PARAMELLE Bernard	Pneumologie	
	PECCOUD François	Analyse (I.U.T. B)	
	PEFFEN René	Métallurgie (I.U.T. A)	
	PERRET Jean	Neurologie	
	PERRIER Guy	Géophysique - Glaciologie	
	PHELIP Xavier	Rhumatologie	
	RACHAIL Michel	Médecine Interne	
	RACINET Claude	Gynécologie et Obstétrique	
	RAMBAUD André	Hygiène et Hydrologie	
	RAMBAUD Pierre	Pédiatrie	
	Mme	RENAUDET Jacqueline	Bactériologie
	MM.	ROBERT Jean-Bernard	Chimie-Physique
ROMIER Guy		Mathématiques (I.U.T. B)	
SHOM Jean-Claude		Chimie générale	
STOEBNER Pierre		Anatomie pathologique	
VROUSOS Constantin		Radiologie	

MAITRE DE CONFERENCES ASSOCIES

M. COLE Antony Sciences Nucléaires

CHARGE DE FONCTIONS DE MAITRE DE CONFERENCES

M. JUNIEN-LAVILLAVROY Paul O.R.L.

Fait à SAINT MARTIN D'HERES,  
DECEMBRE 1975.



Monsieur BARRA me fait l'honneur de présider mon jury de thèse,  
Messieurs GEFFROY et ROMIER d'en être membres. Qu'ils soient assurés de ma  
reconnaissance.

Que tous ceux qui ont contribué à la réalisation de ce travail soient  
ici remerciés. En particulier :

Monsieur VAN CUTSEM qui a patronné avec bienveillance ce travail, en  
m'offrant sa confiance et me faisant profiter de son expérience ;

Madame CORDEL qui a dactylographié avec diligence le manuscrit ;

Monsieur LAPERRIERE qui en a effectué le tirage.

Je remercie, enfin, mes collègues chambériens a côté desquels j'ai  
poursuivi mes recherches.



## TABLE DES MATIERES

Introduction		1
Notations		12
Chapitre I	DESCRIPTION, TABULATION, PROPRIETES DES TESTS FONDAMENTAUX (13 - 37)	
§0	Introduction et notations	13
§1	Principaux tests existants	14
§2	Equivalences	25
§3	Effet de cache	28
§4	Problèmes posés par des applications successives de la règle de Thompson généralisée	34
Chapitre II	REGLES DE BAYES GENERALISEE (39 - 62)	
§0	Introduction et notations	39
§1	Extension du risque de Bayes	40
§2	Caractérisation des règles doubles de Bayes dans le cas d'une fonction de coût $L_0$ simple	41
§3	Caractérisation dans un cas particulier	46
§4	Autre propriété caractéristique	47
§5	Propriétés des règles doubles de Bayes symétri- ques R.P.D.	49
§ 6	Résultats sur l'invariance	51

Chapitre III	TESTS OPTIMAUX POUR PLUSIEURS VALEURS ABERRANTES DE MATRICE DES VARIANCES-COVARIANCES $\Lambda$ CONNUE (63 - 82)	
§0	Introduction et notations	63
§1	Glissements colinéaires	64
§2	Applications	74
§3	Glissements orthogonaux	77
§4	Conclusion	82
Chapitre IV	TESTS OPTIMAUX POUR PLUSIEURS VALEURS ABERRANTES DE MATRICE DES VARIANCES-COVARIANCES $\Lambda$ INCONNUE (83 - 103)	
§0	Introduction et notations	83
§1	Formulation du problème et théorème principal	83
§2	Applications	98
§3	Glissements orthogonaux	99
§4	Conclusion	103
Chapitre V	QUELQUES SUGGESTIONS - QUELQUES EXEMPLES (105 - 123)	
§0	Introduction	105
§1	Traitement d'un échantillon multidimensionnel composante par composante	105
§2	Tests de normalité	107
§3	Echantillon I.N.S.E.E.	107
§4	Deuxième échantillon Algorithme de la règle de Thompson généralisée avec winsorisation généralisée	114 121
§5	Suggestions	123
Bibliographie		125

## INTRODUCTION

Le traitement des valeurs aberrantes, que nous étudions ici, est l'un des plus vieux et des plus constants problèmes de la statistique. Dans un échantillon à valeurs dans  $\mathbb{R}$ , il arrive qu'une ou plusieurs observations se détachent en étant soit plus grandes, soit plus petites que toutes les autres. Nous les appelons valeurs douteuses. Elles peuvent aussi bien

\* être des réalisations de la variable aléatoire  $X$  associée à l'échantillon, se situant dans un intervalle  $]-\infty, a]$  ou  $[b, +\infty[$  de probabilité faible. Elles proviennent alors de la variabilité intrinsèque des données.

\* être le résultat d'une erreur : erreur de manipulation, erreur de mesure due à un appareillage défectueux, erreur de calcul ; ou d'une autre cause connue ou inconnue.

Le statisticien désire introduire un niveau plus ou moins objectif à partir duquel il va considérer ces valeurs comme étrangères à l'échantillon : c'est-à-dire aberrantes. Voici les raisons principales qui motivent cette détection :

a) Epurer les données en vue de leur utilisation statistique : en particulier avant une estimation, problème longuement traité par les américains sous le titre : "Rejection of outliers".

b) Contrôler les simulations, et tester la fiabilité des processus génératifs.

c) Isoler des observations, d'un intérêt particulier justement parce qu'elles sont extrêmes.

Il ne faut donc pas limiter le traitement des valeurs aberrantes à un simple problème de rejet. Voir Anscombe, F.J. (1960) et Kruskal, W.H. (1960).

En dimension 1, valeur douteuse est donc synonyme de valeur extrême.

Dans le cas d'un échantillon à valeurs dans  $\mathbb{R}^p$ , la première difficulté

consiste à définir une valeur douteuse.

Ce peut être non seulement un vecteur dont l'une des composantes est nettement plus grande ou petite que la composante correspondante des autres vecteurs, mais aussi un vecteur dont plusieurs composantes sont entachées d'erreurs moins voyantes.

Le cas le plus généralement étudié est celui d'un échantillon normal gaussien : i.e.  $n$  réalisations  $x_1, \dots, x_n$  supposées indépendantes d'une variable aléatoire  $X$

- à valeurs dans  $\mathbb{R}$  d'espérance  $\mu$  et d'écart type  $\sigma$ , de densité  $\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2(\frac{x-\mu}{\sigma})^2}$
- à valeurs dans  $\mathbb{R}^p$  d'espérance  $\underline{m} \in \mathbb{R}^p$  et de matrice des variances-covariances  $\Lambda$  de densité  $\frac{1}{\sqrt{(2\pi)^p |\Lambda|}} e^{-1/2(\underline{x}-\underline{m})' \Lambda^{-1} (\underline{x}-\underline{m})}$  où  $|\Lambda|$  est le déterminant de  $\Lambda$ .

Pour structurer notre problème nous utilisons deux modèles proposés par Grubbs et Dixon en 1950 en dimension 1, repris par différents auteurs, puis généralisés en dimension  $p$ .

Modèle **A** : une valeur aberrante provient d'un glissement de la moyenne, soit d'une loi  $\mathcal{N}(\mu+\Delta\mu, \sigma^2)$  où  $\Delta\mu \neq 0 \in \mathbb{R}$  [resp. d'une loi  $\mathcal{N}_p(\underline{m}+\Delta\underline{m}, \Lambda)$  où  $\Delta\underline{m}$  est un vecteur non nul de  $\mathbb{R}^p$ ]

Modèle **B** : une valeur aberrante provient d'une variation de la dispersion soit d'une loi  $\mathcal{N}(\mu, \sigma^2\lambda^2)$  où  $\lambda > 1$ . [resp. d'une loi  $\mathcal{N}_p(\underline{m}, \lambda\Lambda)$  où  $\lambda > 1$ ]

Ces modèles nous fournissent une définition probabiliste d'une valeur aberrante et permettent l'étude de la puissance de différents tests. Les valeurs aberrantes donnent lieu principalement aux études suivantes :

- \* élaboration et tabulation des tests ;
- \* étude de puissance intrinsèque et comparative des tests ;
- \* étude des conséquences des rejets sur l'estimation ; étude de techniques particulières : censure, winsorisation, etc...

Nous ne prétendons pas présenter ici de façon exhaustive les différents travaux sur les valeurs aberrantes. Nous renvoyons pour cela au livre de David, M.A. (1970) chapitre 8 : "The treatment of outliers" ainsi qu'à la généreuse bibliographie

de la fin de l'ouvrage. Nous ne signalons que les éléments qui nous paraissent essentiels tant du point de vue heuristique que du cadre particulier de notre problème, que nous délimitons ci-après.

Anscombe, F.J.(1960) prétend que l'une des premières allusions aux valeurs douteuses a été faite par Bessel, le célèbre astronome, qui dans un article publié en 1838, se refuse à éliminer une donnée sous le seul prétexte qu'elle est entachée d'une déviation (écart à la moyenne) importante. "Nous croyons, dit-il, que seule l'observation rigoureuse de cette règle écartera l'arbitraire de nos résultats".

Toujours d'après Anscombe, ce fut Peirce, B.(1852) qui proposa le premier, un critère de rejet, repris et simplifié par Chauvenet, W.(1863). Ce test est basé sur l'hypothèse qu'il n'y a au plus qu'une valeur aberrante et se présente de la façon suivante : posons

$$p = \int_{d_p}^{+\infty} dF(x)$$

où  $dF(x)$  est la densité de la variable  $X$  à valeurs dans  $\mathbb{R}$ .

Soit  $m$  le nombre de réalisations  $x_i$  de  $X$  telles que  $x_i \in ]d_p + \infty[$

Alors  $m$  suit une loi binomiale  $M$  de probabilité.

$$P[M = m] = C_n^m p^m (1-p)^{n-m} \text{ et l'on a } E(M) = np.$$

Chauvenet décida que  $E(M)$  devait toujours valoir  $1/2$ . Pour tout  $n$ , on peut

calculer  $d_p$  tel que

$$\int_{d_p}^{+\infty} dF(x) = \frac{1}{2n}$$

Soit  $x_i$  la valeur douteuse (unique par hypothèse); la règle s'énonce : si  $x_i > d_p$  rejeter  $x_i$ . Chauvenet raisonne en nombre moyen de valeurs non aberrantes et désire que celui-ci vaille  $1/2$ . Habituellement on raisonne de la façon suivante :

l'ensemble critique  $E$  est la réunion des ensembles où l'un des  $x_i$  est  $> d_p$

et les autres inférieurs. La probabilité de  $E$  est alors

$$d = P[E] = n \frac{1}{2n} \left(1 - \frac{1}{2n}\right)^{n-1}$$

ce qui nous donne :

n	1	2	3	4	∞
α	0,500	0,375	0,347	0,335	0,303

En testant un grand nombre d'échantillons on considère, dans environ 1/3 des échantillons, des valeurs comme aberrantes alors qu'elles ne le sont pas. Le principal défaut du test de Chauvenet est son niveau de signification déraisonnable. Après la description du test de Peirce, Chauvenet ajoute : "Le critère ci-dessus repose sur des principes probabilistes qui peuvent paraître obscurs aux non initiés". Ne pouvons-nous pas lui appliquer la même remarque ? Toujours est-il que le test de Peirce, celui de Chauvenet et d'autres suggestions provoquèrent de vives discussions et le sujet devint classique dans les traités sur les moindres carrés, à la fin du siècle dernier.

Après Chauvenet ce fut Wright, T.W. (1884) qui fit le pas le plus marquant. Il suggère de rejeter toute observation dont la déviation dépasse en grandeur 5 fois l'écart probable (écart ayant la probabilité 1/2 d'être dépassé en valeur absolue) soit 3,37 fois l'écart type. Lorsque l'hypothèse de normalité est vérifiée, seule 1 observation sur mille est rejetée. Il en résulte peu de perte d'information.

Cette règle est studentisée par Thompson, W.R. (1935) qui propose de considérer comme aberrante toute observation  $x_i$  telle que  $\left| \frac{x_i - \bar{x}}{s} \right| > c_a$  où :  $c_a$  est une constante qui dépend du niveau de signification,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  et  $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ . Cette règle nous satisfait en apparence puisqu'elle traite le cas de plusieurs valeurs aberrantes éventuelles. Hélas, Pearson, E.S. et Chandra-Sekar, C. (1936) mirent en évidence une carence importante de la règle de Thompson. Pour des échantillons de petite taille et des taux de rejet modérés, il est impossible de rejeter plus d'une observation au moyen de cette règle : ce qui signifie que si par malchance deux observations sont pareillement éloignées de la moyenne, aucune ne sera rejetée. C'est ce qu'on appelle l'effet de Cache (Masking effect). Cependant la règle de Thompson reste valable si la taille de l'échantillon est élevée et le taux

de rejet conséquent.

Grubbs, F.E.(1950), proposa le test associé à  $\frac{x_{(n)} - \bar{x}}{s}$  où

$x_{(n)} \geq x_{(n-1)} \geq \dots \geq x_{(1)}$  désigne l'échantillon ordonné.

Il donne une forme équivalente de ce test, au moyen de l'égalité :

$$\frac{S^2(n)}{S^2} = 1 - \frac{1}{n-1} \frac{(x_{(n)} - \bar{x})^2}{s^2}$$

où

$$S^2 = ns^2$$

$$S^2(n) = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_{(i)} - \bar{x}(n))^2, \quad \bar{x}(n) = \frac{1}{n-1} \sum_{i=1}^{n-1} x_{(i)}$$

Ce test peut largement se généraliser. Pour tester par exemple les deux valeurs

$x_{(n)}$  et  $x_{(n-1)}$  on calcule

$$\frac{S^2(n, n-1)}{S^2} = \frac{\sum_{i=1}^{n-2} (x_{(i)} - \bar{x}(n, n-1))^2}{s^2}$$

avec

$$\bar{x}(n, n-1) = \frac{1}{n-2} \sum_{i=1}^{n-2} x_{(i)}$$

Dixon, W.J.(1950) quant à lui, proposa une classe de statistiques de tests dont la principale est

$$r_{10} = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$$

Si  $r_{10}$  dépasse une certaine valeur critique que l'on trouve dans les tables fournies par Dixon on considère  $x_{(n)}$  comme aberrante. Cette statistique se généralise en posant

$$r_{ij} = \frac{x_{(n)} - x_{(n-i)}}{x_{(n)} - x_{(j+1)}}$$

pour se préserver de l'effet de Cache. Dixon suggère de traiter le cas de plusieurs valeurs douteuses par un emploi répété de  $r_{22}$ , ceci après avoir étudié, par simulation, la performance de ses tests- en utilisant les modèles **A** et **B**.

Les procédures selon lesquelles  $x_{(n)}$  est déclaré aberrante, si  $\frac{x_{(n)} - \bar{x}}{s}$  dépasse une valeur critique, possèdent une propriété d'optimalité étudiée par Murphy, R.B.(1951), Paulson, E.(1952) et Kudo, A.(1956a). Ils envisagent le problème comme un problème de décision multiple et se restreignent au modèle **A** et aux

règles invariantes sous une transformation  $y = ax + b$ ,  $a$  et  $b$  réels.

Murphy suppose qu'il y a  $k$  valeurs aberrantes dans l'échantillon ou 0. Ces valeurs proviennent de loi  $N(\mu + \alpha_i \Delta\mu, \sigma)$  pour  $i = (n - k + 1), \dots, n$ . La procédure qui maximise la probabilité de rejet des  $k$  valeurs aberrantes parmi les procédures invariantes et de même rang ( $P_{H_0}(D_0) = 1 - \alpha$ ) est associée à

$$\alpha_n \frac{(x_{(n)} - \bar{x})}{s} + \dots + \alpha_{n-k+1} \frac{(x_{(n-k+1)} - \bar{x})}{s}$$

où  $\alpha_n \geq \alpha_{n-1} \geq \dots \geq \alpha_{n-k+1}$  sont connues.

En conséquence lorsque les glissements en moyenne sont inconnus, il n'existe pas de règle possédant la même propriété d'optimalité.

Paulson considère un échantillon réparti en  $r$  classes de  $k$  valeurs soit au total  $rk = n$  observations. Il se demande si l'une des classes a glissé en moyenne et montre que le test possédant la propriété d'optimalité analogue à celle du test de Murphy est associée à

$$\frac{\bar{x}_{(r)} - \bar{x}}{s} \quad \text{avec} \quad \bar{x}_{(r)} = \max_{\ell=1, \dots, r} \bar{x}_{\ell}$$

où  $\bar{x}_{\ell} = \frac{1}{k} \sum_{j=1}^k x_{k(j-1)+1}$  est la moyenne de la  $\ell^{\circ}$  classe,  $\bar{x}$  et  $s$  étant respectivement la moyenne et l'écart type de l'échantillon global.

Kudo répartit ses observations en 3 groupes indépendants. Le premier suit une loi  $\mathcal{N}(\mu, \sigma^2)$  sauf, peut être, pour une valeur, ayant glissé en moyenne. Le second suit une loi  $\mathcal{N}(\mu, \sigma^2)$  et ne contient aucune valeur aberrante. Le troisième suit une loi  $\mathcal{N}(\mu_1, \sigma_1^2)$  où  $\mu_1 \neq \mu$ . Lorsqu'il n'y a pas d'observations dans le second groupe, le test proposé par Kudo est celui pour lequel Quesenberry, C.P. et David, H.A. (1961) ont calculé les valeurs critiques. Lorsque seul le 1<sup>o</sup> groupe contient des observations, le test se réduit à  $\frac{x_{(n)} - \bar{x}}{s}$ .

Beaucoup d'autres travaux ont été effectués en dimension 1. Nous citerons Mac Millan, R.G. (1968) qui étudie la puissance de trois tests pour deux valeurs aberrantes. Il s'agit

1° du test séquentiel associé à  $\frac{x_{(n)} - \bar{x}}{s}$  ;

2° du test de Murphy où  $\alpha_1 = \alpha_2 = 1$  soit  $\frac{x_{(n)} + x_{(n-1)} - 2\bar{x}}{s}$  ;

3° du test de Grubbs lorsque les 2 plus grandes observations sont douteuses, soit  $\frac{S^2(n, n-1)}{s^2}$

Plus récemment, il semble que l'on marque un intérêt croissant pour les valeurs aberrantes dans les modèles linéaires avec un système d'erreurs indépendantes, normales et homoscédatiques.

Le premier test concernant les valeurs douteuses dans un échantillon normal gaussien multidimensionnel de matrice des variances-covariances  $\Lambda$  quelconque mais supposée connue, a été proposé par Kudô, A. (1957). Il est associé à la statistique

$$\chi^2(n) = \max_{\mu=1 \dots n} (x_{\mu} - \bar{x})' \Lambda^{-1} (x_{\mu} - \bar{x}) .$$

Kudô démontre ensuite une propriété d'optimalité de ce test. Une propriété analogue fut démontrée, lorsque  $\Lambda$  est inconnue et estimée par

$S = \frac{1}{n-1} \sum_{\mu=1}^n (x_{\mu} - \bar{x})(x_{\mu} - \bar{x})'$  par Karlin, S. et Truax, D.R. (1960) dans le cas du modèle  $\mathcal{A}$  et par Ferguson, T.S. (1961b) dans le cas du modèle  $\mathcal{B}$ . Le test est alors associé à

$$T^2(n) = \max_{\mu=1 \dots n} {}^t(x_{\mu} - \bar{x}) S^{-1} (x_{\mu} - \bar{x}) .$$

Siotani, M. (1959) a étudié la distribution asymptotique de  $\chi^2(n)$ .

Enfin Wilks, S.S. (1963) a proposé un test général basé sur

$$r_{\xi} = \det \frac{(a_{ij\xi})}{(a_{ij})}$$

où  $(a_{ij}) = (n-1) S$  et  $a_{ij\xi} = \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n (x_{\mu i} - \bar{x}_{\xi i})(x_{\mu j} - \bar{x}_{\xi j})$

avec  $\bar{x}_{\xi i} = \frac{1}{n-1} \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n x_{\mu i}$

Si la quantité  $r_{(1)} = \min_{\xi=1\dots n} r_{\xi}$  est inférieure à une valeur critique que l'on trouve dans les tables fournies par Wilks, on considère l'observation correspondante comme aberrante.

Ce test se généralise au cas de plusieurs valeurs douteuses (cf Chapitre 1).

Beaucoup de travail reste à faire sur les valeurs aberrantes multidimensionnelles. Nous nous sommes donc intéressés aux échantillons gaussiens à valeurs dans  $\mathbb{R}^p$ , contenant éventuellement plusieurs valeurs douteuses. Nous nous sommes situés dans le cadre du modèle  $\mathcal{L}$  où l'on suppose que les valeurs aberrantes proviennent de glissements en moyenne.

Dans un souci de clarté, nous présentons au Chapitre I les différents tests multidimensionnels existants, leurs propriétés et leur tabulation. Nous démontrons en particulier l'équivalence du test de Karlin et Truax-Ferguson avec le test de Wilks pour une valeur aberrante ; puis l'équivalence du même test avec le test associé à

$$Q_{\xi}^2(n) = \max_{\xi=1\dots n} \frac{n-1}{n} (x_{\xi} - \bar{x}(\xi))' U^{-1}(\xi) (x_{\xi} - \bar{x}(\xi))$$

où  $\bar{x}(\xi)$  et  $U(\xi)$  sont respectivement la moyenne et la matrice des variances-covariances estimées à partir de l'échantillon privé de la  $\xi^{\circ}$  observation. Puis nous mettons en évidence un effet identique à celui étudié par Pearson et Chandra-Sekar en dimension 1. Dans le cas d'un échantillon multidimensionnel de taille modérée, et d'un pourcentage de rejet faible, la règle de Thompson généralisée qui consiste à rejeter toute valeur  $x_{\xi}$  pour laquelle  $T_{\xi}^2$  dépasse une valeur critique, dépendant du niveau de signification, ne rejette au plus qu'une seule valeur. Au cas où plusieurs valeurs sont également éloignées de la moyenne  $\bar{x}$  au sens de la distance généralisée, cette règle n'en rejette aucune. A fortiori le test de Karlin et Truax-Ferguson souffre de cet effet.

Nous soulevons ensuite un problème posé par des applications successives d'un même test sur un échantillon, en ôtant à chaque passage les valeurs détectées lors du passage précédent, et en conservant le même niveau de signification ; ceci jusqu'à l'arrêt de la détection.

On s'aperçoit qu'une telle méthode risque de nous priver d'une partie importante des observations.

Dans le chapitre II, nous présentons une généralisation des règles de Bayes dans le cas à nous avons  $m$  hypothèses alternatives :  $H_1 \dots H_m$  et des densités dépendant d'un paramètre  $\Delta m \in \mathbb{R}^k$ . Nous identifions alors l'ensemble des états de la nature:  $\mathcal{E}$  au produit  $\{ H_0, H_1, \dots, H_m \} \times \mathbb{R}^k$  sur lequel nous définissons la probabilité à-priori  $b \otimes \eta$  où  $b$  est une probabilité discrète a priori sur  $\{H_0, \dots, H_m\}$  et  $\eta$  une probabilité a priori absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^k$ .

Nous définissons ensuite la fonction de coût  $L_0$  qui nous servira tout au long de notre étude et obtenons une caractérisation des règles doubles de Bayes que nous précisons dans le cas d'une probabilité  $b$  particulière. Nous obtenons alors une propriété caractéristique des règles doubles de Bayes sous forme d'équivalence, que nous utilisons pour démontrer une propriété des règles doubles de Bayes symétriques relativement aux probabilités de décision.

Enfin nous démontrons au dernier paragraphe du Chapitre II quelques résultats sur l'invariance. En particulier nous obtenons un théorème qui nous permet de nous limiter, lors de la recherche des règles doubles de Bayes parmi les règles invariantes sous deux groupes de transformation, à l'étude de l'invariant maximal sous ces deux groupes.

Puis nous donnons au chapitre III un modèle général de valeurs aberrantes provenant de glissements en moyenne proportionnels. Nous traitons ce problème comme un problème de décision multiple et cherchons la procédure qui, parmi les procédures symétriques, invariantes par translation, de même niveau  $P_{H_0}(D_0) = 1 - \alpha$ , maximise  $P_{H_{\nu}, \Delta m}^{\nu}(D_{\nu})$  uniformément en  $\nu = 1, \dots, m$ . Ce qui revient à rechercher une règle double de Bayes, selon notre terminologie. Comme corollaire, dans le cas où il y a  $k$  valeurs aberrantes ou 0, de loi respectives.  $d_{\mu}^{\nu}(m + \alpha_{\mu} \Delta m, \Lambda)$  pour  $\mu = 1 \dots k$ ,  $\alpha_{\mu} \neq 0$ ,  $\Delta m \in \mathbb{R}^p$ ,  $\Lambda$  étant supposée connue, la procédure optimale est associée à

$$R^2_{(\sigma)} = \max_{\sigma \in \mathfrak{G}} \left[ \sum_{\mu=1}^k \alpha_{\mu} (x_{\sigma(\mu)} - \bar{x}) \right]' \Lambda^{-1} \left[ \sum_{\mu=1}^k \alpha_{\mu} (x_{\sigma(\mu)} - \bar{x}) \right]$$

où  $\mathfrak{G}$  est le groupe des permutations de l'ensemble  $\{1, \dots, n\}$

Nous obtenons également comme corollaire le résultat suivant : si les observations sont réparties en  $r$  classes de  $k$  éléments, la procédure optimale qui détecte si l'une des classes a glissé en moyenne est associée à

$$R^2_{(\ell)} = \max_{1 \leq \ell \leq r} (\bar{x}(\ell) - \bar{x})' \Lambda^{-1} (\bar{x}(\ell) - \bar{x})$$

où  $\bar{x}(\ell)$  est la moyenne de la  $\ell^{\circ}$  classe.

Nous avons supposé ensuite qu'il y avait deux valeurs aberrantes dans l'échantillon ou 0, ces deux valeurs provenant de glissements en moyenne orthogonaux au sens de l'inverse de la matrice des variances-covariances. La procédure optimale est alors associée à

$\chi^2_{(n)} + \chi^2_{(n-1)}$  ( $n$ ) et ( $n-1$ ) étant les 2 premiers indices pour lesquels on obtient les deux plus grandes valeurs de  $\chi^2_{\mu}$ .

Nous étudions les mêmes problèmes, au chapitre IV, dans le cas où  $\Lambda$  est inconnue. Nous sommes amenés à nous limiter aux règles de décision invariantes non seulement sous le groupe des translations, mais aussi sous le groupe linéaire  $GL(p, \mathbb{R})$ . Nous trouvons alors que la règle qui maximise la probabilité de rejet de  $k$  valeurs aberrantes de lois respectives définies comme ci-dessus, mais avec  $\Lambda$  inconnue cette fois, est associée à la statistique

$$T^2_{(\sigma)} = \max_{\sigma \in \mathfrak{G}} \left[ \sum_{\mu=1}^k \alpha_{\mu} (x_{\sigma(\mu)} - \bar{x}) \right]' S^{-1} \left[ \sum_{\mu=1}^k \alpha_{\mu} (x_{\sigma(\mu)} - \bar{x}) \right]$$

où  $S$  est la matrice estimée sans biais de  $\Lambda$ . Ce qui est la généralisation en dimension  $p$  du résultat de Murphy, R.B. (1951) en dimension 1.

Enfin lorsque nous supposons qu'il y a deux valeurs aberrantes ou 0, provenant de glissements orthogonaux, la règle optimale est associée à la statistique

$$T^2_{(n)} + T^2_{(n-1)}$$

Nous signalons que ce résultat se généralise aisément au cas de  $k > 2$  valeurs aberrantes provenant de glissements deux à deux orthogonaux au sens du produit scalaire associé à l'inverse de la matrice des variances-covariances.

Dans le chapitre V, nous décrivons et étudions deux échantillons. Le premier nous est fourni par les résultats du premier tour des élections présidentielles de 1974 de MM. GISCARD d'ESTAING et MITTERAND. Le second, à valeurs dans  $\mathbb{R}^3$ , nous a été fourni par l'INSEE. Il est constitué de 252 entreprises suivies sur 11 années (de 1963 à 1973) pour lesquelles on connaît le logarithme des 3 paramètres suivants : travail, capital, production.

Nous proposons pour résoudre un problème mis en lumière au chapitre I une méthode que nous appelons winsorisation généralisée.

Enfin nous envisageons certains prolongements de notre étude.



## NOTATIONS

Le transposé d'un vecteur  $X$  sera noté  $X'$ . De même pour une matrice.

$\underline{x} = x_1 \dots x_n$  désigne un échantillon gaussien multidimensionnel

$n$  : taille de l'échantillon

$p$  : dimension de l'échantillon

$m$  : nombre d'hypothèses alternatives

$\mu, \xi$  numérotent les vecteurs

$i, j$  numérotent les composantes de ces vecteurs

$v$  numérote les hypothèses alternatives

$\mathcal{H}$  : ensemble d'hypothèses

$\mathcal{E} = \mathcal{H} \times \mathbb{R}^k$

$\mathcal{D}$  : ensemble des décisions

$\mathbb{D}$  : ensemble des règles de décision aléatoire

$d$  : règle de décision

$m$  : vecteur moyenne

$\bar{x}$  : vecteur moyenne estimé

$\Delta m$  : variation du vecteur moyenne

$\Lambda$  : matrice des variances-covariances

$S$  : matrice estimée des variances-covariances

$b$  : probabilité a priori sur  $\mathcal{H}$

$\eta$  : probabilité a priori sur  $\mathbb{R}^k$

$A \otimes B$  : produit de Kronecker (ou tensoriel) de deux matrices

$|B| = \det B$

$L$  : fonction de coût



## CHAPITRE I

### §0 INTRODUCTION ET NOTATIONS

Nous présentons ici les principaux tests multidimensionnels concernant les valeurs aberrantes dans un échantillon gaussien à valeurs dans  $\mathbb{R}^p$  de loi  $\mathcal{N}_p(m, \Lambda)$ . Nous étudions certaines relations entre ces tests et leurs propriétés. Nous mettons en évidence un effet parasite dû à la présence de plusieurs valeurs aberrantes dans l'échantillon. Enfin nous envisageons une application séquentielle de l'un des tests et soulevons le problème de l'arrêt de la méthode.

Nous précisons ici quelques notations :

$n$  désigne la taille de l'échantillon,  $p$  sa dimension.

Nous supposons toujours  $n > p > 1$

Les lettres grecques numérotent les vecteurs de  $\mathbb{R}^p$ .

Les lettres latines, les composantes de ces vecteurs.

La matrice des variances-covariances sera généralement désignée par  $\Lambda$  et l'espérance dans  $\mathbb{R}^p$  par  $m$ .

Soit donc  $x_1, \dots, x_n$  un échantillon gaussien, i.e.  $n$  réalisations indépendantes d'une loi de probabilité gaussienne  $\mathcal{N}_p(m, \Lambda)$  à valeurs dans  $\mathbb{R}^p$ .

Posons

$$\bar{x} = \frac{1}{n} \sum_{\mu=1}^n x_{\mu}$$

$$S = \frac{1}{n-1} \sum_{\mu=1}^n (x_{\mu} - \bar{x})(x_{\mu} - \bar{x})'$$

$\bar{x}$  et  $S$  sont les estimés sans biais respectifs de la moyenne  $m$  et de la matrice des variances-covariances  $\Lambda$ , associés à l'échantillon.

## §1 LES PRINCIPAUX TESTS EXISTANTS

### I.1. La règle de Kudo.

Ce fut Kudo, A. (1957) qui le premier proposa une règle de détection d'une valeur aberrante multidimensionnelle. Ayant constaté qu'en statistique d'ordre : valeur douteuse et valeur extrême sont synonymes, il propose de généraliser la notion de valeur extrême en dimension  $p > 1$ , à l'aide de l'écart généralisé studentisé à la valeur moyenne. Une valeur extrême est une valeur pour laquelle cet écart est maximal.

La règle de Kudo concerne le cas où la moyenne  $m$  est inconnue et la matrice des variances-covariances  $\Lambda$  connue.

Posons

$$(I.2.) \quad \left| \begin{array}{l} \chi^2_{\mu} = (x_{\mu} - \bar{x})' \Lambda^{-1} (x_{\mu} - \bar{x}) \\ \chi^2_{(n)} = \max_{\mu=1 \dots n} \chi^2_{\mu} \end{array} \right.$$

Si  $\chi^2_{(n)}$  dépasse une certaine valeur critique qui dépend du niveau de signification choisi, on considère l'observation correspondante comme aberrante.

Pratiquement ceci revient à déterminer un ellipsoïde de confiance centré sur la moyenne estimée  $\bar{x}$ . Si l'observation associée à  $\chi^2_{(n)}$  se situe à l'extérieur de cet ellipsoïde, on la considère comme aberrante.

Kudo en proposant ce test, en a démontré une propriété d'optimalité. Son hypothèse de base est qu'il n'y a au plus qu'une valeur aberrante, associée au modèle  $A$ . (glissement en moyenne).

Il assimile son test à un problème de décision multiple, et démontre que la règle de décision décrite ci-dessus maximise les probabilités de décision uniformément en  $\mu = 1, \dots, n$ , parmi les règles de décision vérifiant certaines propriétés, entre autre d'invariance et de symétrie.

Comme nous obtenons ce résultat au corollaire III.9. comme cas particulier, nous ne décrirons pas ici plus longuement cette propriété.

La tabulation de la statistique correspondante à la règle de Kudo a été

faite par Siotani, M. (1959) pour  $p = 2, 3, 4$  ;  $n = 3$  à  $10$  , puis de deux en deux jusqu'à  $20$  puis  $25$  et  $30$  ;  $\alpha = 5\%$  ,  $2,5\%$  et  $1\%$  .

### 1.3. La règle de Karlin et Truax.

Cette règle est l'analogie de la règle de Kudo, lorsque cette fois, outre la moyenne, la matrice des variances-covariances nous est inconnue.

Soit  $S$  la matrice estimée sans biais des variances-covariances

Nous poserons

$$(I.4.) \quad T_{\mu}^2 = (x_{\mu} - \bar{x})' S^{-1} (x_{\mu} - \bar{x})$$

$$T_{(n)}^2 = \max_{\mu=1 \dots n} T_{\mu}^2$$

Si  $T_{(n)}^2$  dépasse une certaine valeur critique qui dépend du niveau de signification choisi, on considère l'observation correspondante comme aberrante. En fait Karlin et Truax ont envisagé le cas suivant plus général : soit  $\mathcal{S}$  une matrice de Wishart  $p \times p$  de matrice des variances-covariances  $\Lambda$  , qu'ils ont supposé indépendante de  $(X_1, \dots, X_n)$ .  $\mathcal{Y}$  est un estimé, basé sur  $\zeta$  degrés de liberté, de la matrice des variances-covariances  $\Lambda$  , obtenu à l'aide d'un échantillon distinct. Karlin et Truax ont proposé le test associé à la statistique :

$$(I.5.) \quad \max_{\mu=1 \dots n} (x_{\mu} - \bar{x})' (\mathcal{Y} + \underline{xx}' - n \bar{x} \bar{x}')^{-1} (x_{\mu} - \bar{x})$$

où  $\underline{x} = (x_1, \dots, x_n)$  ,  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$

Lorsque  $\mathcal{Y} = 0$  on retrouve le test associé à la statistique (I.4.).

Karlin, S. et Truax, D.R. (1960) ont également envisagé le cas où l'on dispose d'une variable contrôle.

Ils ont ensuite démontré la propriété analogue, lorsque  $\Lambda$  est inconnue, à la propriété d'optimalité de la règle de Kudo.

Cette propriété, Kudo l'avait pressentie, mais n'était pas parvenu à la démontrer. Les hypothèses de base sont les mêmes :

- le problème est un problème de décision multiple
- il n'y a au plus qu'une valeur aberrante dans l'échantillon, associée

au modèle  $A$

- les règles de décision vérifient certaines propriétés d'invariance (une de plus que pour le cas  $\Lambda$  connue).

Karlin et Truax utilisent un théorème qui leur fournit la densité de l'invariant maximal, lorsque la  $\mu^\circ$  observation a glissé en moyenne, grâce à une intégration par rapport à la mesure de Haar sur le groupe linéaire  $GL(p, \mathbb{R})$ .

Nous obtenons le résultat de Karlin et Truax lorsque  $\mathcal{Y} = 0$  comme corollaire au chapitre IV et renvoyons à ce chapitre pour une description complète de cette propriété d'optimalité. Signalons cependant que nous n'utilisons pas le résultat sur le calcul de la densité de l'invariant maximal par intégration par rapport à la mesure de Haar sur le groupe linéaire.

Signalons que pour  $p = 2, n = 3(1) 12, 14$   $a = 0,05, 0,025$  et  $0,01$  Siotani a tabulé la statistique

$$\max_{\mu=1\dots n} (x_\mu - \bar{x})' \mathcal{Y}^{-1} (x_\mu - \bar{x})$$

où  $\mathcal{Y}$  est un estimé sans biais basé sur  $\zeta$  degrés de liberté de  $\Lambda$ , indépendant des  $\underline{x}$ , avec  $\zeta = 20(2) 40(5)60, 100, 150$  et  $200$ . En définitive, nous ne disposons pas directement de tables associées à la statistique I.4.

### I.6. Les tests de Wilks, S.S. (1963).

Ces tests ont le mérite de s'attaquer à plusieurs valeurs aberrantes éventuelles dans l'échantillon et présentent un intérêt particulier dans le cas d'échantillons de taille modérée et de niveau de rejet faible. Voir §3 Effet de cache.

Wilks suppose que la moyenne  $m$  et la matrice des variances-covariances  $\Lambda$  sont inconnues.

Posons

$$a_{ij} = \sum_{\xi=1}^n (x_{\xi i} - \bar{x}_i)(x_{\xi j} - \bar{x}_j) \text{ pour } i, j = 1, \dots, p$$

$$(a_{ij}) = (n-1)S$$

Considérons le nuage de points de  $\mathbb{R}^p$  associé à l'échantillon. Prenons  $p$  de ces  $n$  points auxquels on ajoute le centre de gravité  $\bar{x}$  de l'échantillon.

Ces  $p+1$  points forment un simplexe et Wilks a démontré que la somme des carrés des volumes de tous les simplexes obtenus à partir de l'échantillon est égal à

$$\frac{\det(a_{ij})}{(p!)^2} = \frac{|a_{ij}|}{(p!)^2} .$$

Il appelle  $|a_{ij}|$  la dispersion interne de l'échantillon. Si  $n > p$ ,  $|a_{ij}| > 0$  avec la probabilité 1. En laissant de côté le  $\xi^o$  élément du nuage on obtient de la même façon la dispersion interne de  $n-1$  points restants soit :

$$a_{ij\xi} = \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n (x_{\mu i} - \bar{x}_{\xi i})(x_{\mu j} - \bar{x}_{\xi j}) \quad \text{où} \quad \bar{x}_{\xi i} = \frac{1}{n-1} \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n x_{\mu i}$$

posons

$$(I.7.) \quad r_{\xi} = \frac{|a_{ij\xi}|}{|a_{ij}|} , \quad |a_{ij\xi}| > 0 \quad \text{avec la probabilité 1 si } n > p + 1 .$$

La quantité  $r_{\xi}$  est le rapport de dispersion dans le cas d'une valeur aberrante. Soient  $r_{(1)} \dots r_{(n)}$  Ces quantités rangées par ordre croissant. Le critère de Wilks consiste à regarder si  $r_{(1)}$  est trop petit.

Cette méthode se généralise aisément au cas de  $\ell > 1$  valeurs douteuses. Elle consiste à calculer le rapport des dispersions de l'échantillon privé de  $\ell$  de ses observations et de l'échantillon total, puis à vérifier à l'aide de tables si le plus petit d'entre eux est inférieur ou non à une valeur critique.

Nous explicitons ici les formules lorsque  $\ell = 2$ .

$$a_{ij\xi\theta} = \sum_{\substack{\mu=1 \\ \mu \neq \xi, \theta}}^n (x_{\mu i} - \bar{x}_{(\xi\theta)i})(x_{\mu j} - \bar{x}_{(\xi\theta)j})$$

où

$$\bar{x}_{(\xi\theta)i} = \frac{1}{n-2} \sum_{\substack{\mu=1 \\ \mu \neq \xi, \theta}}^n x_{\mu i}$$

Alors

$$(I.8.) \quad r_{\xi\theta} = \frac{|a_{ij\xi\theta}|}{|a_{ij}|}$$

où  $|a_{ij\xi\theta}|$  est le déterminant de la matrice  $p \times p : (a_{ij\xi\theta})$ , qui est p.s.  $> 0$  si  $n > p + 2$ .

Le test s'opère de façon analogue au cas  $\ell = 1$ . Lorsque  $\ell = 1$  ce test possède la propriété d'optimalité du test de Karlin et Truax (voir §2 Equivalences). Lorsque  $\ell > 1$  ses propriétés n'ont pas été investiguées, sans doute à cause du travail considérable que cela demande.

Wilks a proposé une tabulation approximative de ses tests en utilisant la première inégalité de Bonferroni

$$P[r_{(1)} < r] \leq nP[r_{\xi} < r]$$

à l'aide de la fonction Beta.

Cette tabulation est faite pour  $\ell = 1, 2$

$$n = 5(1) 30.(5) 100(100) 500, p = 1(1) 5$$

$$\alpha = 0,1, 0.05, 0.025, 0.01$$

### 1.9. La règle de Thompson généralisée.

Posons

$$T_{\mu}^2 = (x_{\mu} - \bar{x})' S^{-1} (x_{\mu} - \bar{x})$$

La règle de Thompson généralisée consiste à considérer comme aberrante toute valeur pour laquelle  $T_{\mu}^2$ , appelé écart généralisé studentisé à la moyenne dépasse une valeur critique  $K_{\alpha}$  qui dépend du niveau de signification  $\alpha$ . Ce test très général permet apparemment de rejeter plusieurs valeurs aberrantes éventuelles. Nous mettrons en évidence aux §3 et 4, deux particularités de cette règle qui sont en fait deux inconvénients. Pour sa tabulation nous avons besoin d'une égalité.

Considérons la matrice  $U(\xi)$  définie par

$$(I.10.) \quad U(\xi) = \frac{1}{n-2} \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n (x_{\mu} - \bar{x}(\xi)) (x_{\mu} - \bar{x}(\xi))'$$

où  $\bar{x}(\xi) = \frac{1}{n-1} \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n x_{\mu}$

$U(\xi)$  est la matrice estimée sans biais de la matrice des variances-covariances de l'échantillon privé de la  $\xi^{\circ}$  observation.

Posons alors

$$G_{\xi}^2 = \frac{n-1}{n} (x_{\xi} - \bar{x}(\xi))' U^{-1}(\xi) (x_{\xi} - \bar{x}(\xi))$$

I.11. Proposition.

$$T_{\xi}^2 = \frac{n-1}{n} \frac{(n-1) \mathcal{L}_{\xi}^2}{n-2 + \mathcal{L}_{\xi}^2}$$

(Preuve : pour démontrer cette égalité nous utilisons plusieurs lemmes

Remarquons tout d'abord qu'en posant

$$A = \sum_{\mu=1}^n (x_{\mu} - \bar{x})(x_{\mu} - \bar{x})'$$

$$A(\xi) = \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n (x_{\mu} - \bar{x}(\xi))(x_{\mu} - \bar{x}(\xi))'$$

on a  $A = (n-1)S$  et  $A(\xi) = (n-2) U(\xi)$

I.12. Lemme.

$$\bar{x} = \bar{x}(\xi) + \frac{1}{n} (x_{\xi} - \bar{x}(\xi))$$

(Preuve :

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{\mu=1}^n x_{\mu} = \frac{1}{n} x_{\xi} + \frac{1}{n} \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n x_{\mu} \\ &= \frac{1}{n} x_{\xi} + \frac{n-1}{n} \bar{x}(\xi) = \bar{x}(\xi) + \frac{1}{n} (x_{\xi} - \bar{x}(\xi)). \end{aligned}$$

I.13. Corollaire.

$$x_{\xi} - \bar{x} = \frac{n-1}{n} (x_{\xi} - \bar{x}(\xi))$$

(Preuve :

$$x_{\xi} - \bar{x} = x_{\xi} - \bar{x}(\xi) - \frac{1}{n} (x_{\xi} - \bar{x}(\xi)) = \frac{n-1}{n} (x_{\xi} - \bar{x}(\xi)).$$

I.14. Lemme.

$$A = A(\xi) + \frac{n-1}{n} Y_{\xi} Y_{\xi}' \quad \text{où} \quad Y_{\xi} = x_{\xi} - \bar{x}(\xi)$$

(Preuve : Soit  $a_{ij}$  (resp.  $a_{ij\xi}$ ) le terme de la  $i^{\circ}$  ligne et de la  $j^{\circ}$  colonne de la matrice  $A$  (resp.  $A(\xi)$ ). Il nous faut vérifier que

$$a_{ij} = a_{ij\xi} + \frac{n-1}{n} Y_{\xi i} Y_{\xi j}$$

on a

$$a_{ij} = \sum_{\mu=1}^n (x_{\mu i} - \bar{x}_i)(x_{\mu j} - \bar{x}_j). \text{ En utilisant le lemme 1}$$

$$a_{ij} = \sum_{\mu=1}^n [x_{\mu i} - \bar{x}_i(\xi) - \frac{1}{n}(x_{\xi i} - \bar{x}_i(\xi))] [x_{\mu j} - \bar{x}_j(\xi) - \frac{1}{n}(x_{\xi j} - \bar{x}_j(\xi))]$$

$$= \sum_{\mu=1}^n (x_{\mu i} - \bar{x}_i(\xi))(x_{\mu j} - \bar{x}_j(\xi)) - \frac{1}{n}(x_{\xi i} - \bar{x}_i(\xi)) \sum_{\mu=1}^n (x_{\mu j} - \bar{x}_j(\xi))$$

$$- \frac{1}{n}(x_{\xi j} - \bar{x}_j(\xi)) \sum_{\mu=1}^n (x_{\mu i} - \bar{x}_i(\xi)) + \frac{1}{n}(x_{\xi i} - \bar{x}_i(\xi))(x_{\xi j} - \bar{x}_j(\xi))$$

$$= a_{ij\xi} + \frac{n-1}{n}(x_{\xi i} - \bar{x}_i(\xi))(x_{\xi j} - \bar{x}_j(\xi)) - \frac{1}{n}(x_{\xi i} - \bar{x}_i(\xi)) \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n (x_{\mu i} - \bar{x}_i(\xi))$$

$$- \frac{1}{n}(x_{\xi i} - \bar{x}_i(\xi))(x_{\xi j} - \bar{x}_j(\xi)) - \frac{1}{n}(x_{\xi j} - \bar{x}_j(\xi)) \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n (x_{\mu i} - \bar{x}_i(\xi))$$

$$- \frac{1}{n}(x_{\xi i} - \bar{x}_i(\xi))(x_{\xi j} - \bar{x}_j(\xi))$$

$$\text{or } \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n (x_{\mu i} - \bar{x}_i(\xi)) = 0 = \sum_{\substack{\mu=1 \\ \mu \neq \xi}}^n (x_{\mu j} - \bar{x}_j(\xi))$$

$$\text{d'où } a_{ij} = a_{ij\xi} + \frac{n-1}{n}(x_{\xi i} - \bar{x}_i(\xi))(x_{\xi j} - \bar{x}_j(\xi))$$

$$= a_{ij\xi} + \frac{n-1}{n} Y_{\xi i} Y_{\xi j} \quad \text{c.q.f.d.)}$$

### I.15. Lemme.

Soient  $A$  une matrice réelle  $p \times p$ , symétrique  $\lambda$  un scalaire réel et  $Y$  un vecteur de  $\mathbb{R}^p$ . On a la relation

$$Y' \tilde{A}' Y = Y' A + \lambda Y Y'$$

où pour une matrice  $A$ ,  $\tilde{A}$  désigne sa matrice des cofacteurs.

(Preuve : la matrice  $A$  étant symétrique  $\tilde{A}' = \tilde{A}$ . De même  $\lambda YY'$  est une matrice symétrique ; a fortiori la matrice  $A + \lambda YY'$ . Par conséquent, on a également

$$\widetilde{[A + \lambda YY']'} = \widetilde{A + \lambda YY'}$$

Nous allons donc montrer que pour tout vecteur  $Y$  de  $\mathbb{R}^p$  et tout scalaire  $\lambda \in \mathbb{R}$

$$\widetilde{A} Y = \widetilde{A + \lambda YY'} Y$$

Le lemme I.15. en résultera. Dans une base  $A$  représente l'endomorphisme  $f$  et  $\lambda YY'$  l'endomorphisme  $\varphi$ . Nous noterons, pour une matrice  $A$  donnée, d'endomorphisme  $f$  associé,  $f^*$  l'endomorphisme associé à  $\tilde{A}$ . On veut donc montrer que  $\forall Y \in \mathbb{R}^p, \lambda \in \mathbb{R}$

$$(f + \varphi)^*(Y) = f^*(Y)$$

On sait que si  $\omega$  désigne une forme  $p$ -linéaire alternée non nulle sur  $\mathbb{R}^p$  (par exemple si  $e_1 \dots e_p$  désignent les  $p$  vecteurs de la base canonique prendre  $e_1 \wedge e_2 \wedge \dots \wedge e_p$ ) et si  $\theta$  désigne un endomorphisme de  $\mathbb{R}^p$  et  $Z, x_1 \dots x_{p-1}$ ,  $p$  vecteurs de  $\mathbb{R}^p$  :

$$\begin{aligned} & \omega(\theta^*(Z), x_1, \dots, x_{p-1}) \\ &= \omega(Z, \theta(x_1), \dots, \theta(x_{p-1})) \end{aligned}$$

soit

$$\omega((f + \varphi)^*(Y), x_1, \dots, x_{p-1}) = \omega(Y, (f + \varphi)(x_1), \dots, (f + \varphi)(x_{p-1}))$$

Calculons le membre de droite.

$\omega$  étant multilinéaire, cette expression vaut

$$\omega(Y, f(x_1), \dots, f(x_{p-1})) + \Delta$$

où  $\Delta$  est formé de termes du type  $\omega(Y, \xi_1, \dots, \xi_{p-1})$ ,  $\xi_1, \dots, \xi_{p-1}$  étant soit des  $f(x_i)$  soit des  $\varphi(x_i)$   $i = 1 \dots p-1$ , et l'un au moins des  $\xi_i$  vaut  $\varphi(x_i)$ .

Mais

$$\varphi(x_i) = \lambda YY' x_i = \alpha_i Y \text{ où } \alpha_i = \lambda Y' x_i \in \mathbb{R}$$

On en déduit que chaque terme de  $\Delta$  est du type :

$$\alpha \omega(Y, \eta_1, \dots, \eta_{p-1})$$

où  $\alpha$  est un produit de  $\alpha_i$  et les  $\eta_i$  sont soit des  $f(x_i)$  soit des  $Y$  l'un au moins étant un  $Y$ .

Mais alors deux des vecteurs sur lesquels opère  $\omega$  sont identiques : chaque terme composant  $\Delta$  est donc nul soit

$$\Delta = 0$$

$$\text{Donc } \omega((f + \gamma)^*(Y), x_1, \dots, x_{p-1}) = \omega(f^*(Y), x_1, \dots, x_{p-1})$$

les vecteurs  $x_1 \dots x_{p-1}$  étant quelconques dans  $\mathbb{R}^p$  et  $\omega$  régulière donc

$$(f + \gamma)^*(Y) = f^*(Y) \text{ soit } \widetilde{A + \lambda Y Y' Y} = \widetilde{A Y}$$

I.16. Lemme.

Soient  $x$  et  $y$  deux vecteurs de  $\mathbb{R}^p$ . Notons  $I$  la matrice identité de  $\mathbb{R}^p$  et  $\langle \dots \rangle$  le produit scalaire canonique.

Alors  $\det(I + yx') = 1 + \langle x, y \rangle$

(Preuve : il existe une transformation orthogonale  $V$  telle que

$$x = Vx_1 \text{ où } x_1 \text{ est parallèle au premier vecteur de base } e_1 = \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}.$$

Alors

$$\begin{aligned} \det(I + yx') &= \det(I + y x_1' V^{-1}) \\ &= \det(I + V^{-1} y x_1') = \det(I + V^{-1} y x_1') \end{aligned}$$

Posons

$$V^{-1}y = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}, \text{ alors}$$

$$V^{-1} y x_1' = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} (x_1, 0, \dots, 0) = \begin{pmatrix} a_1 & x_1 & 0 & \dots & 0 \\ a_2 & x_1 & 0 & & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ a_p & x_1 & 0 & \dots & 0 \end{pmatrix}$$

La matrice  $I + V^{-1} y x_1'$  est une matrice triangulaire inférieure dont le déterminant est égal au produit des termes diagonaux.

$$\begin{aligned} \det(I + V^{-1} y x_1') &= 1 + a_1 x_1 = 1 + \langle V^{-1}y, V^{-1}x \rangle \\ &= 1 + \langle y, x \rangle \\ &= 1 + \langle x, y \rangle \end{aligned}$$

car la transformation orthogonale  $V$  conserve le produit scalaire canonique. )

A l'aide de ces 5 résultats démontrons la proposition I.11.

$$T_{\xi}^2 = (x_{\xi} - \bar{x}) S^{-1} (x_{\xi} - \bar{x}) = \frac{(n-1)^2}{n^2} (x_{\xi} - \bar{x}(\xi))' S^{-1} (x_{\xi} - \bar{x}(\xi)),$$

ceci d'après le corollaire I.13.. Donc, en tenant compte de nos notations

$$T_{\xi}^2 = \frac{(n-1)^3}{n^2} Y_{\xi}' A^{-1} Y_{\xi}$$

En utilisant le lemme I.14.

$$\begin{aligned} T_{\xi}^2 &= \frac{(n-1)^3}{n^2} Y_{\xi}' [A(\xi) + \frac{n-1}{n} Y_{\xi} Y_{\xi}' ]^{-1} Y_{\xi} \\ &= \frac{(n-1)^3}{n^2} Y_{\xi}' \frac{A(\xi) + \frac{n-1}{n} Y_{\xi} Y_{\xi}'}{\det [A(\xi) + \frac{n-1}{n} Y_{\xi} Y_{\xi}']} Y_{\xi} \end{aligned}$$

or 
$$A(\xi) + \frac{n-1}{n} Y_{\xi} Y_{\xi}' = A(\xi) [I + \frac{n-1}{n} A^{-1}(\xi) Y_{\xi} Y_{\xi}']$$

Appliquons le lemme I.16. à

$$x = \frac{n-1}{n} A^{-1}(\xi) Y_{\xi} \quad y = Y_{\xi}$$

Il vient

$$\det[A(\xi) + \frac{n-1}{n} Y_{\xi} Y_{\xi}'] = \det A(\xi) (I + \frac{n-1}{n} Y_{\xi}' A^{-1}(\xi) Y_{\xi})$$

d'où

$$T_{\xi}^2 = \frac{(n-1)^3}{n^2} \frac{Y_{\xi}' A(\xi) + \frac{n-1}{n} Y_{\xi} Y_{\xi}' Y_{\xi}}{\det A(\xi) (1 + \frac{n-1}{n} Y_{\xi}' A^{-1}(\xi) Y_{\xi})}$$

Utilisons le lemme I.15.

$$T_{\xi}^2 = \frac{(n-1)^3}{n^2} \frac{Y_{\xi}' A(\xi) Y_{\xi}}{\det A(\xi) (1 + \frac{n-1}{n} Y_{\xi}' A^{-1}(\xi) Y_{\xi})}$$

En multipliant numérateur et dénominateur par  $(n-2)$  il vient

$$\begin{aligned} T_{\xi}^2 &= \frac{n-1}{n} \frac{(n-1) \frac{n-1}{n} Y_{\xi}' U^{-1}(\xi) Y_{\xi}}{n-2 + \frac{n-1}{n} Y_{\xi}' U^{-1}(\xi) Y_{\xi}} \\ &= \frac{n-1}{n} \frac{(n-1) \mathcal{C}_{\xi}^2}{n-2 + \mathcal{C}_{\xi}^2} \end{aligned}$$

c.q.f.d.)

Tabulation de  $T_{\xi}^2$  : posons  $v = \mathcal{C}_{\xi}^2$ ,  $v > 0$  et étudions la fonction

$$f(v) = \frac{(n-1)v}{n-2+v} > 0 \text{ pour } n > 2$$

Soit  $f'$  la dérivée de  $f$

$$f'(v) = \frac{(n-1)(n-2)}{(n-2+v)^2} > 0$$

Donc  $T_{\xi}^2$  est une fonction strictement croissante de  $\mathcal{C}_{\xi}^2$  et

$$T_{\xi}^2 > k \iff \mathcal{C}_{\xi}^2 > \frac{n(n-2)k}{(n-1)^2 - (n-1)k}$$

Nous savons (voir par exemple Anderson, T.W. (1958)) que  $\mathcal{C}_{\xi}^2$  suit une loi de Hotelling à  $n-2$  degrés de liberté et que  $(n-p)\mathcal{C}_{\xi}^2 / (n-1)p$  suit une loi  $F$  à  $p$  et  $n-p$  degrés de liberté.

Pour un niveau de signification  $\alpha$ , fixé, la valeur critique nous est donnée par

$$K_{\alpha} = \frac{(n-1)^2 F_{p, n-p, \alpha}}{\frac{n(n-p)(n-2)}{(n-1)p} + nF_{p, n-p, \alpha}}$$

### I.17. Travaux récents.

Pour terminer cette présentation des tests multidimensionnels, de détection de valeurs aberrantes, signalons les articles de Gnanadesikan, R. et Kettenring, J.R. (1972) et de Hawkins, D.M. (1974). Les tests proposés par ce dernier supposent implicitement la connaissance de  $m$  et de  $\Lambda$  ce qui est rarement le cas, ou du moins d'un premier estimé  $\bar{x}$  de  $m$  et d'un second estimé  $S$  de  $\Lambda$  indépendants de l'observation testée. Le test est alors associé à la statistique.

$$T_1 = (x_0 - \bar{x})' S^{-1} (x_0 - \bar{x})$$

où  $x_0$  est l'observation indépendante de  $S$  et  $\bar{x}$ .

Pour la tabulation de ce test on sait que

$$\frac{(n-p)n}{(n-1)(n+1)p} T_1 \text{ suit une loi } F(p, n-p) \text{ (voir Anderson, T.W. (1958) pages 107 et 123)}$$

La nouveauté des autres tests provient du fait que les auteurs étudient le problème des valeurs aberrantes sous l'angle d'une analyse en composantes principales. Pour plus de détails, nous renvoyons aux articles correspondants.

## §2 EQUIVALENCES

Nous voulons établir ici certaines relations entre les tests explicités au §1.

### I.18. Lemme.

On a la relation :

$$r_{\xi} = 1 - \frac{n}{(n-1)^2} T_{\xi}^2 \quad \text{pour } \xi = 1 \dots n$$

(Preuve :

$$r_{\xi} = \frac{|a_{ij\xi}|}{|a_{ij}|}$$

Posons 
$$b_{\xi i} = \sqrt{\frac{n}{n-1}} (x_{\xi i} - \bar{x}_i)$$

alors 
$$a_{ij\xi} = a_{ij} - b_{\xi i} b_{\xi j} \quad \text{d'où}$$

$$\begin{aligned} |a_{ij\xi}| &= |a_{ij} - b_{\xi i} b_{\xi j}| = \det [(a_{ij}) - b_{\xi} \cdot b'_{\xi}] \\ &= \det(a_{ij}) \det [I_p - (a_{ij})^{-1} b_{\xi} b'_{\xi}] \end{aligned}$$

D'après le lemme I.16.

$$\det [I_p - (a_{ij})^{-1} b_{\xi} b'_{\xi}] = 1 - \langle b_{\xi}, (a_{ij})^{-1} b_{\xi} \rangle$$

or 
$$(a_{ij})^{-1} = \frac{1}{n-1} S^{-1}, \text{ donc}$$

$$\begin{aligned} r_{\xi} &= 1 - \langle \sqrt{\frac{n}{n-1}}(x_{\xi i} - \bar{x}_i), \frac{1}{n-1} S^{-1} \sqrt{\frac{n}{n-1}}(x_{\xi i} - \bar{x}_i) \rangle \\ &= 1 - \frac{n}{(n-1)^2} T_{\xi}^2 \quad .) \end{aligned}$$

### I.19. Proposition.

La statistique du test de Wilks pour une valeur aberrante est équivalente à celle du test de Karlin et Truax.

(Preuve : pour  $u > 0$  la fonction  $f(u) = 1 - \frac{n}{(n-1)^2} u^2$  est une fonc-

tion croissante de  $u$  .

$r_{\xi}$  est donc fonction croissante de  $T_{\xi}^2$

d'où

$$\boxed{r(1) = 1 - \frac{n}{(n-1)^2} T^2(n)} \quad . )$$

I.20. Proposition.

La statistique du test de Karlin et Truax est équivalente à la statistique

$$T_{\xi}^2(n) = \max_{\xi=1 \dots n} T_{\xi}^2$$

où  $T_{\xi}^2$  est le réel défini en I.10.

(Preuve : A la proposition I.11. nous avons démontré l'égalité

$$T_{\xi}^2 = \frac{(n-1)^2 T_{\xi}^2}{n(n-2) + n T_{\xi}^2}$$

Considérons la fonction  $f(u) = \frac{(n-1)^2 u^2}{n(n-2) + nu^2}$

$$f'(u) = \frac{2(n-1)^2(n-2)u}{n(n-2+u^2)^2} > 0 \quad \text{si } u > 0$$

(I.21.) Rappel : nous avons supposé  $n > p \geq 2$  .

Donc pour  $u > 0$   $f$  est fonction croissante de  $u$  .  $T_{\xi}^2$  est une fonction croissante de  $T_{\xi}^2$  et l'on a

$$T^2(n) = \frac{(n-1)^2 T_{\xi}^2(n)}{n(n-2) + n T_{\xi}^2(n)} \quad . )$$

I.22. Quelques remarques sur ces équivalences.

Nous savons que le test de Karlin et Truax possède la double propriété d'être à la fois optimal sous le modèle  $\mathcal{H}$  et sous le modèle  $\mathcal{B}$  (voir Karlin, S. et Truax, D.R. (1960) ainsi que Ferguson, T.S. (1961, b)). Ces équivalences nous permettent de donner une justification intuitive de ces résultats.

Précisons tout d'abord ce que sont les deux modèles  $\mathcal{A}$  et  $\mathcal{B}$ . Le modèle  $\mathcal{A}$  suppose qu'une valeur étrangère à l'échantillon est une réalisation d'une loi  $\omega_p(m + \Delta m, \Lambda)$ . La valeur étrangère provient donc d'une loi dont les ellipsoïdes de confiance sont déduits des ellipsoïdes de confiance de la loi  $\omega_p(m, \Lambda)$  par une simple translation. Le modèle  $\mathcal{B}$  suppose qu'une valeur aberrante est une réalisation d'une loi  $\omega_p(m, \lambda\Lambda)$  avec  $\lambda > 1$ . Dans ce cas, les ellipsoïdes associés à cette loi restent centrés en  $m$  mais se déduisent des ellipsoïdes de confiance de loi  $\omega_p(m, \Lambda)$  par une dilatation de rapport  $\lambda$  au sens de la distance généralisée. En présentant son test, Wilks nous en a fourni un point de vue géométrique décrit en I.6.

Son test est associé au rapport de deux sommes de carrés de volumes de simplexes : la première obtenue en privant l'échantillon de la  $\xi^\circ$  observation, la seconde obtenue à partir de l'échantillon global.

$$r_\xi = \frac{|a_{ij\xi}|}{|a_{ij}|} = \frac{\frac{|a_{ij\xi}|}{(p!)^2}}{\frac{|a_{ij}|}{(p!)^2}}$$

Si la  $\mu^\circ$  observation est nettement (et anormalement) éloignée du reste des observations, les volumes des simplexes formés à partir de ce point seront considérablement grands. La valeur  $|a_{ij\mu}|$  sera, elle, plus petite que les autres et le rapport  $r_\mu$  sera le plus petit des  $n$  rapports  $r_\xi$  : c'est l'observation associée qui sera éventuellement détectée.

D'un autre point de vue, si l'on regarde  $\mathcal{C}_\xi^2$ , c'est à un coefficient près, la distance généralisée de la  $\xi^\circ$  observation à la valeur moyenne  $\bar{x}(\xi)$  de l'échantillon privée de cette observation. Au cas où  $x_\xi$  est une valeur aberrante, les estimés  $\bar{x}(\xi)$  et  $U(\xi)$  ne seront pas entachés des biais consécutifs à la valeur  $x_\xi$  : la mise en évidence de  $x_\xi$  comme valeur aberrante à partir de  $\mathcal{C}_\xi^2$  n'en sera que plus nette. Intuitivement ce test a toute chance d'être performant dans le cadre des deux modèles. C'est à ce test que le test de Karlin et Truax est équivalent.

### §3 EFFET DE CACHE

Mis en évidence en dimension 1 par Pearson, E.S. et Chandra Sekar, C.(1936), il a été étudié par Murphy, R.B.(1951) et Mac Millan, R.G.(1968) qui proposèrent tous deux des tests dans le cas de plusieurs valeurs aberrantes (en dimension 1 toujours)

Pearson et Chandra Sekar ont en effet montré que pour des échantillons de petites tailles et des niveaux de rejet peu élevés ( $\alpha = 0,05$ , ou  $\alpha = 0,01$ ) la règle de Thompson associée à la statistique

$$\frac{|x_{\mu} - \bar{x}|}{s} \quad \text{où } \bar{x} = \frac{1}{n} \sum_{\mu=1}^n x_{\mu} \quad \text{et } s^2 = \frac{1}{n} \sum_{\mu=1}^n (x_{\mu} - \bar{x})^2$$

ne détecte au plus qu'une valeur à la fois. Ce qui veut dire que si par malchance deux observations sont pareillement et anormalement éloignées de la valeur moyenne cette règle n'en détectera aucune. C'est ce qu'on appelle l'effet de cache (Masking effect, en anglais).

La présence de plusieurs valeurs aberrantes dans un échantillon biaise considérablement la moyenne et l'écart type estimés, ce dernier en l'augmentant, ce qui réduit leur déviation : elles se cachent alors les unes les autres.

Nous montrons ici sur un exemple que cet inconvénient semble persister en dimension  $p > 1$ , pour des échantillons de petite taille ( $n < 20$ ) et des niveaux de confiance élevés (99 %) avec la règle I.9. et a fortiori avec la règle I.3.

Nous avons tiré au hasard un sous-échantillon de taille  $n = 14$ , de l'échantillon gaussien bidimensionnel décrit au chapitre V. Nous noterons cet échantillon  $Ech_0$ . Il nous est fourni par le tableau I.23.

Nous lui avons appliqué la règle de Thompson généralisée. (règle I.9) avec un niveau de confiance de 99 % soit  $\alpha = 0,01$ . Le tableau I.24. nous donne les écarts généralisés studentisés correspondants. La valeur critique étant égale à 6,71 nous détectons la valeur (- 13,88 ; 1,44).

A l'aide de cet échantillon nous avons procédé à environ une trentaine de passages-machine en lui ajoutant une ou deux valeurs "fantômes" délibérément étrangères

TABLEAU I.23

P=2 et N=14

VALERY GISCARD D'ESTAING	FRANCOIS MITTERAND	DEPARTEMENTS NUMEROTES	
10.90	-9.99	ILLE-ET-VILLAIN	35
-0.60	-0.76	HAUTES-ALPES	5
-7.78	11.73	SEINE-SAINT-DENIS	93
3.12	-1.57	YONNE	89
-13.88	1.41	LANDES	40
-1.75	3.89	ISERE	38
2.82	1.77	HAUTE-SAONE	70
-6.20	13.77	NIEVRE	58
7.76	-14.10	HAUT-RHIN	68
5.19	-1.40	DOUBS	25
-5.12	3.82	LOT	46
-1.35	-2.75	LOIR-ET-CHER	41
6.00	-0.82	MOSELLE	57
-1.38	4.49	VAL-D'OISE	95

TABLEAU I.24

ECARTS GENERALISES EN ORDRE DECROISSANT

40	LANDES	7.9937
68	HAUT-RHIN	4.2802
58	NIEVRE	3.5476
35	ILLE-ET-VILLAIN	2.8907
93	SEINE-SAINT-DENIS	2.3143
57	MOSELLE	1.2627
25	DOUBS	0.8043
41	LOIR-ET-CHER	0.7641
70	HAUTE-SAONE	0.6473
46	LOT	0.5662
95	VAL-D'OISE	0.3513
89	YONNE	0.2432
38	ISERE	0.2079
5	HAUTES-ALPES	0.1266

Total des écarts généralisés TG = 13\*2 = 25.9999

à l'échantillon, que nous avons fait varier à chaque passage. Nous obtenons donc des échantillons de 15 ou 16 éléments.

Le tableau I.25. nous décrit certains passages-machine. Suivant les cas: zéro ou, au plus, une valeur est détectée.

Tableau I.25.

taille	nature de l'échantillon	valeur critique pour $\alpha=0,01$	nombre de valeurs détectées	écart de la valeur détectée ou écart le plus grand (suivant les cas)
14	Ech <sub>0</sub>	6,71	1	7,99
15	Ech <sub>1</sub> = Ech <sub>0</sub> U{(20;20)}	6,87	1	10,48
15	Ech <sub>2</sub> = Ech <sub>0</sub> U{(-13,5;1,2)}	6,87	0	4,84
15	Ech <sub>3</sub> = Ech <sub>0</sub> U{(15; 15)}	6,87	1	9,06
16	Ech <sub>4</sub> = Ech <sub>1</sub> U{(20; 20)}	7,02	0	5,80
16	Ech <sub>5</sub> = Ech <sub>3</sub> U{(14; 14)}	7,02	0	5,66
16	Ech <sub>6</sub> = Ech <sub>3</sub> U{(-15; -15)}	7,02	0	6,43
	etc...			

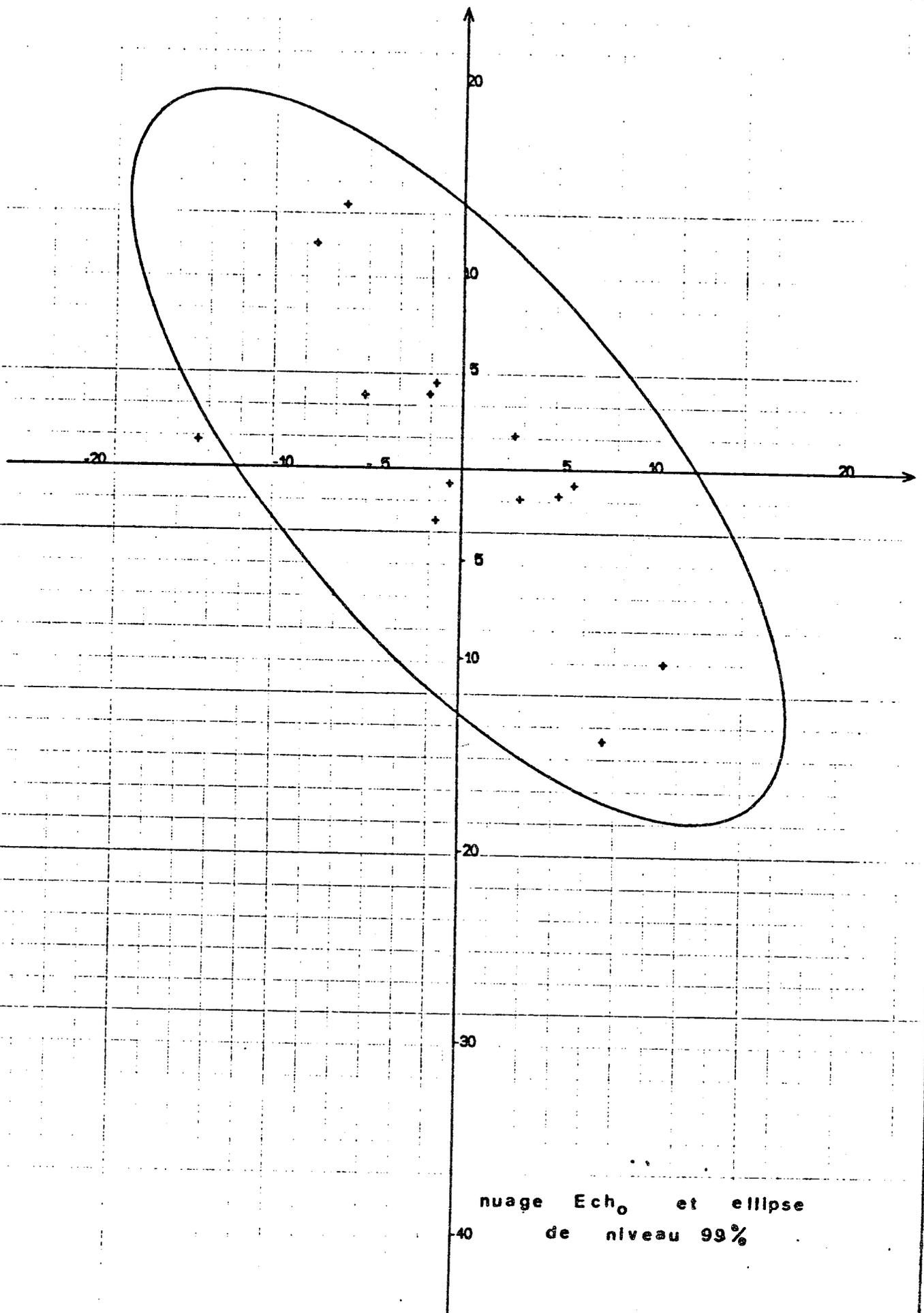
On constate donc sur cet exemple le problème évoqué plus haut. Il est très vraisemblable qu'en dimension  $p > 2$ , ce problème se maintienne. On a en effet le résultat :

I.26. Proposition.

Soit  $x_1, \dots, x_n$  un  $n$  échantillon gaussien multidimensionnel à valeurs dans  $\mathbb{R}^p$ . Soient  $T_1^2, \dots, T_n^2$  les déviations studentisées correspondantes.

Alors

$$\sum_{\mu=1}^n T_{\mu}^2 = (n-1) p$$



(Preuve :

$$\begin{aligned} T_{\mu}^2 &= (x_{\mu} - \bar{x})' S^{-1} (x_{\mu} - \bar{x}) \\ &= \sum_{i,j=1}^p s^{ij} (x_{\mu i} - \bar{x}_i)(x_{\mu j} - \bar{x}_j) \end{aligned}$$

en notant

$$\begin{aligned} S^{-1} &= (s^{ij}) \\ \sum_{\mu=1}^n T_{\mu}^2 &= \sum_{\mu=1}^n \sum_{i,j} s^{ij} (x_{\mu i} - \bar{x}_i)(x_{\mu j} - \bar{x}_j) \\ &= \sum_{i,j} s^{ij} \sum_{\mu=1}^n (x_{\mu i} - \bar{x}_i)(x_{\mu j} - \bar{x}_j) \\ &= \sum_{i,j} s^{ij} (n-1) s_{ij} \end{aligned}$$

en notant

$$S = (s_{ij})$$

$$\sum_{\mu=1}^n T_{\mu}^2 = (n-1) \text{Trace}({}^t S^{-1} S)$$

or

$${}^t S^{-1} = S^{-1} \quad \text{d'où}$$

$$\sum_{\mu=1}^n T_{\mu}^2 = (n-1) \text{trace} (S^{-1} S) = (n-1)p.$$

p étant fixé, n également, la somme des déviations généralisées studentisées est donc une constante qui ne dépend pas de l'échantillon. On se trouve donc dans le cas suivant :

$$\begin{aligned} T_{\mu}^2 &\geq 0 \\ \sum_{\mu=1}^n T_{\mu}^2 &= \sum_{\mu=1}^n T_{(\mu)}^2 = (n-1)p \end{aligned}$$

où

$$T_{(1)}^2 \leq T_{(2)}^2 \leq \dots \leq T_{(n)}^2$$

représentant ces déviations ordonnées.

On voit rapidement que  $T_{(n-1)}^2$  ne pourra en aucun cas dépasser  $\frac{(n-1)p}{2}$  (c'est le cas où

$$T_{(n)}^2 = T_{(n-1)}^2 = \frac{(n-1)p}{2} \quad T_{(\mu)}^2 = 0 \quad \text{pour } i = 1, \dots, n-2)$$

et  $T_{(n-i)}$  ne pourra jamais dépasser  $\frac{(n-1)p}{\mu+1}$   
 $(T_{(n)}^2 = T_{(n-1)}^2 = \dots = T_{(n-\mu)}^2 = \frac{(n-1)p}{\mu+1}$ , les autres tous nuls), ces cas étant évidemment des cas extrêmes jamais réalisés. La somme de ces déviations étant constante, lorsque plusieurs valeurs aberrantes sont présentes, elles se "répartissent" si bien cette constante qu'il n'est plus possible de détecter aucune d'entre elles.

Cette remarque pourrait nous inciter à proposer une tabulation exacte de  $T_{(n)}^2$  : en effet pour  $u > \frac{(n-1)p}{2}$

$$P[T_{(n)}^2 > u] = P([T_1^2 > u] \cup T_2^2 > u] \cup \dots \cup [T_n^2 > u])$$

Soit en désignant par  $T^2$  la variable qui suit la loi commune aux  $T_\mu^2$   
 $\mu = 1 \dots n$ ,  $= \sum_{\mu=1}^n P[T^2 > u] = n P [ T^2 > u ]$ .

Etant donné qu'il est possible de connaître  $P [T^2 > u]$  pour différentes valeurs de  $u$ , on obtiendrait une tabulation exacte de  $T_{(n)}^2$ . Malheureusement de l'identité I.19

$$r_{(1)} = 1 - \frac{n}{(n-1)^2} T_{(n)}^2$$

on déduit

$$T_{(n)}^2 = \frac{(n-1)^2}{n} (1 - r_{(1)})$$

Or nous savons (voir Wilks, S.S. (1963)) que pour  $n > p+1$ ,  $r_\xi$  est positif avec la probabilité 1 ; en donnant à  $r_{(1)}$  la valeur 0 on trouve

$$T_{(n)}^2 = \frac{(n-1)}{n}$$

d'où  $P [T_{(n)}^2 > \frac{(n-1)^2}{n}] = 0$

Comme pour  $p \geq 2$

$$\frac{(n-1)^2}{n} < \frac{(n-1)p}{2}$$

, la tabulation proposée ne fournira jamais la valeur critique exacte.

Néanmoins la distribution de  $T_{(n)}^2$  semble extrêmement difficile à tabuler. Cependant nous proposons pour tabuler  $T_{(n)}^2$  de prendre pour valeur critique de niveau  $\alpha$  la borne supérieure  $u_\alpha$  telle que

I.27.  $P[T_{(n)}^2 > u_\alpha] \leq n P [T^2 > u_\alpha] = \alpha$

Nous signalons de plus le fait intéressant suivant.

I.28. Propriété de  $u$

Considérons les variables aléatoires  $Z_\mu$  telle que

$$Z_\mu(u) = \begin{cases} 1 & \text{si } T_\mu^2 > u \\ 0 & \text{si } T_\mu^2 \leq u \end{cases}$$

de  $\mathbb{R}^+$  dans  $\{0, 1\}$

Calculons alors

$$d(u) = \sum_{\mu=1}^n Z_\mu(u)$$

$d(u)$  représente le nombre de valeurs  $T_\mu^2$  de l'échantillon qui ont dépassé  $u$ .

$$\begin{aligned} E(d(u)) &= E\left(\sum_{\mu=1}^n Z_\mu(u)\right) = E(Z_1(u)) + \dots + E(Z_n(u)) \\ &= n P[T^2 > u] \end{aligned}$$

Donc le nombre moyen de déviations  $T_1^2 \dots T_n^2$  dépassant la valeur  $u$  est égal à la borne supérieure  $n P[T^2 > u]$  de  $P[T_{(n)}^2 > u]$ .

En particulier si  $u_\alpha$  désigne la valeur  $u$  telle que  $n P[T^2 > u_\alpha] = \alpha$ , on a

$$E(d(u_\alpha)) = \alpha$$

I.29. Premiers pas de la tabulation

1° Les valeurs  $u_\alpha$  peuvent se déduire des valeurs  $r_\alpha$  indiquées par Wilks (1963) pour la tabulation de

$$r_{(1)} = \min_{\xi=1 \dots n} \frac{\det[(n-2) S(\xi)]}{\det(n-1) S}$$

via l'identité

$$T_{(n)}^2 = \frac{(n-1)^2}{n} (1 - r_{(1)})$$

2° Pour  $\alpha'$  fixé nous avons vu en I.9 comment trouver la valeur  $u_{\alpha'}$  telle que

$$P[T_\xi^2 > u_{\alpha'}] = \alpha'$$

On en déduit

$$u_{\alpha} : \begin{cases} u_{\alpha} = u_{\alpha'} \\ \alpha = n\alpha' \end{cases}$$

Le tableau ci-dessous permet de faire une comparaison sommaire des valeurs  $u_{\alpha}$  obtenues à partir de la tabulation de Wilks avec celles obtenues par la méthode que nous proposons.

$$\alpha = 0,10 \quad , \quad n = 10$$

Tableau I.30.

p	$u_{\alpha}$ (Wilks)	$u_{\alpha}$ (2° méthode)
2	5,93	5,74
3	6,72	6,50
4	7,30	7,07
5	7,70	7,49

#### §4 PROBLEMES POSES PAR DES APPLICATIONS SUCCESSIVES

##### DE LA REGLE DE THOMPSON GENERALISEE

A la suite d'une détection résultant d'une première application de l'un des différents tests décrits dans la première partie de ce chapitre, on est naturellement tenté d'appliquer une seconde fois le test sur l'échantillon restant (i.e. privé de la valeur ou des valeurs détectées) pour voir si une ou plusieurs autres valeurs sont détectées. Dans son livre, Order Statistics, David, H.A. (1970) envisage ce problème et suggère, lorsqu'on soupçonne plusieurs valeurs, une application répétée du même test, en ajustant bien entendu, la valeur critique  $\lambda_{\alpha}$  en fonction de la taille de l'échantillon ; ceci jusqu'à l'arrêt complet des détections.

Les tests proposés par Kudo, Karlin et Truax, Ferguson et le premier test de Wilks ne détectent qu'une seule valeur à la fois. Or la taille des échantillons est sans cesse croissante, ceci en partie grâce aux possibilités offertes par un traitement informatique des données. Ces règles risquent donc de nécessiter de nombreux

passages et d'élever le coût de l'opération.

C'est pourquoi nous proposons d'appliquer la règle de Thompson généralisée pour traiter les gros échantillons ; cette règle, rappelons-le, a le mérite de détecter plusieurs valeurs à la fois.

Or en appliquant la méthode suggérée par David, on s'aperçoit qu'un grand nombre de valeurs risquent d'être détectées et éventuellement rejetées.

Dans l'exemple traité dont le dessin I.31. nous donne une représentation, nous disposons d'un échantillon gaussien bidimensionnel de moyenne et de matrice des variances-covariances inconnues, de taille 105. Cet échantillon est décrit et étudié en détail au chapitre V. Nous avons procédé sur cet échantillon à une application répétée de la règle de Thompson généralisée I.9. avec un niveau de rejet  $\alpha = 0,05$ .

Le programme calcule dans l'ordre :

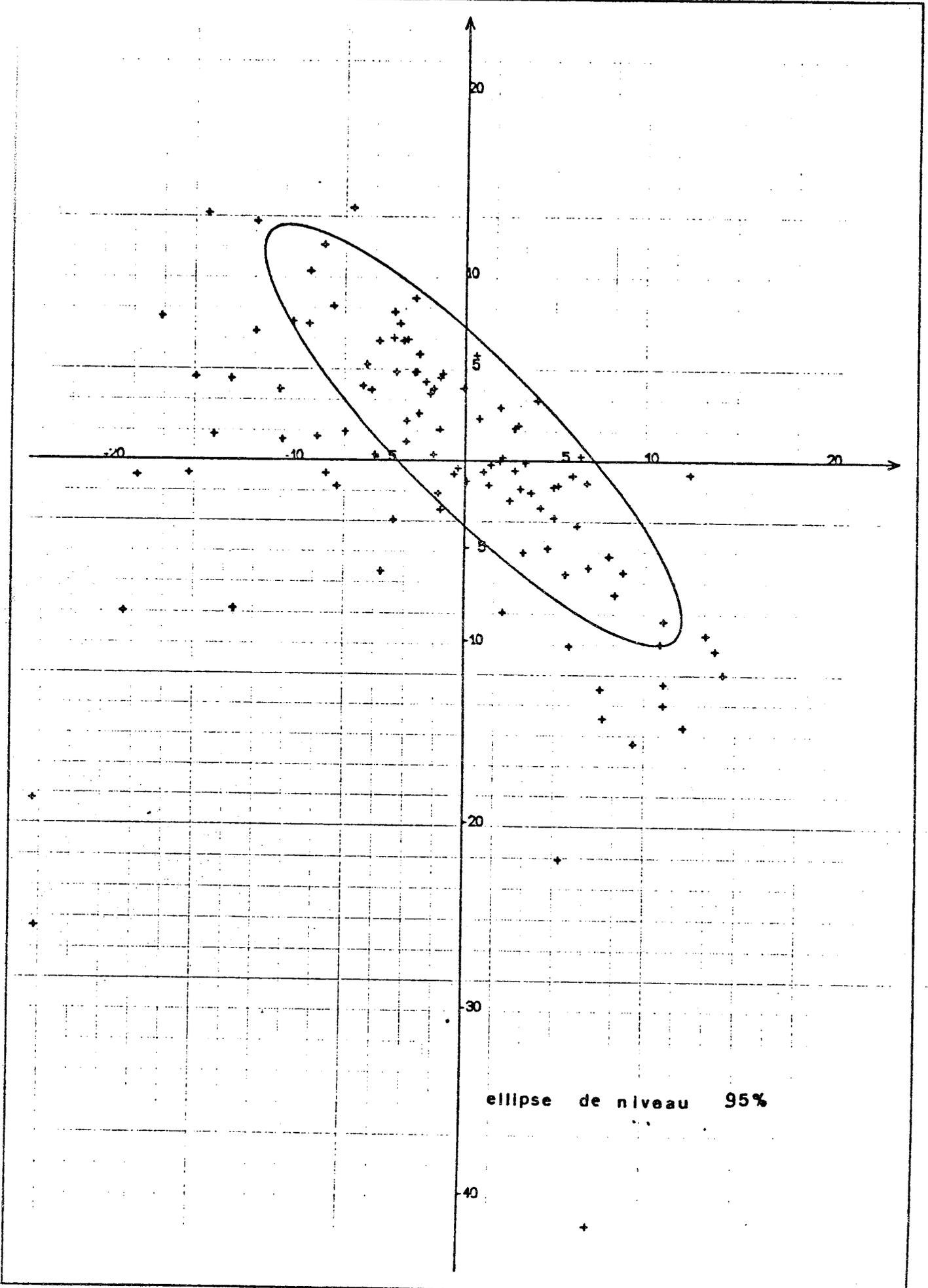
- les composantes du vecteur moyenne  $\bar{x}$
- la matrice des variances-covariances estimée sans biais  $S$
- le déterminant  $DS$  de  $S$
- l'inverse  $S^{-1}$  de la matrice  $S$
- les écarts généralisés studentisés  $T_{\mu}^2$  en ordre croissant
- la valeur critique qui dépend de  $n$  et de  $\alpha$
- la somme des écarts généralisés studentisés.

Cette somme est une constante dont la valeur théorique vaut  $(n-1)p$ ,  $n$  étant la taille de l'échantillon, et,  $p$  sa dimension.

Cette somme a été calculée pour servir de vérification à nos calculs.

A la suite de chaque passage le programme trie l'échantillon et ne garde que les valeurs non détectées (dont l'écart généralisé est inférieur à la valeur critique).

On s'est arrêté au bout de 13 passages, aucune valeur n'étant alors détectée. Mais lors de ces 13 passages, 38 valeurs avaient été successivement détectées, ne nous en laissant que 67, i.e. nous privant de près de 40 % des observations.



ellipse de niveau 95%

L'ellipse tracée est l'ellipse de niveau de confiance 95 % pour les 67 observations restantes, c'est-à-dire l'ensemble des

$$\underline{x} = (x_1, x_2) \quad \text{tels que}$$

$$(x - \bar{x}(67))' S^{-1}(67)(x - \bar{x}(67)) = 5,81$$

$\bar{x}(67)$  et  $S^{-1}(67)$  étant respectivement la moyenne et la matrice des variances-covariances estimées à l'aide des 67 observations restantes.

Nous nous contentons pour le moment de soulever ce problème et renvoyons le lecteur au chapitre V - où il trouvera une étude et une description complète de l'échantillon - pour y discuter la solution proposée.

## CHAPITRE II

### REGLES DE BAYES

#### Introduction et notations

Nous voulons donner ici une justification théorique générale, dans le cadre Bayésien, de la méthode que nous employons pour démontrer l'optimalité des tests, présentés au chapitre III et IV, dans le cas d'un échantillon gaussien multidimensionnel de taille  $n$  et de matrice des variances-covariances  $\Lambda$ . Notre problème est un problème de décision multiple dont nous présentons ici les principaux éléments.

1° Soit  $\mathcal{E}$  l'ensemble des états de la nature. Nous identifions  $\mathcal{E}$  à l'espace produit  $\mathcal{H} \times \mathbb{R}^k$  où  $\mathcal{H}$  est l'ensemble discret des  $m+1$  hypothèses :

$\mathcal{H} = \{H_0, H_1, \dots, H_m\}$  et où  $k$  est la dimension de l'espace d'un paramètre réel.  $\mathbb{R}^k$  est muni de sa tribu des boréliens,  $\mathcal{H}$  de son ensemble des parties  $\mathcal{P}(\mathcal{H})$  et  $\mathcal{E}$  de la tribu produit.

2° Nous disposons sur  $\mathcal{E}$  d'une probabilité a-priori que nous notons  $b \otimes \eta$  où  $b$  est une probabilité discrète "a priori" sur  $\mathcal{H}$  et  $\eta$  une probabilité "a priori" absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^k$  : Nous noterons  $b[H_\nu] = b_\nu$  la probabilité "a priori" de l'hypothèse  $H_\nu$   $\nu = 0, 1, \dots, m$ . On a donc  $\sum_{\nu=0}^m b_\nu = 1$

3° Nous noterons  $\mathcal{D}$  l'ensemble discret des décisions possibles et de même cardinal que  $\mathcal{H}$ .  $\mathcal{D} = \{D_0, D_1, \dots, D_m\}$  où  $D_\nu$  représente la décision d'accepter l'hypothèse  $H_\nu$ , pour  $\nu = 0, 1, \dots, m$ .

4° Nous possédons  $n$  réalisations indépendantes  $x_1, \dots, x_n$  d'une variable aléatoire  $X$  à valeurs dans  $(\mathbb{R}^p)$ , absolument continue par rapport à la mesure de Lebesgue sur  $(\mathbb{R}^p)$ . Nous noterons  $f_\nu(x_1, \dots, x_n)$  la densité conjointe des  $x_\mu$  pour  $\mu = 1, \dots, n$  sous l'hypothèse  $H_\nu$ , pour  $\nu = 0, 1, \dots, m$ . De plus les  $f_\nu$  pour  $\nu = 0, 1, \dots, m$  dépendent d'un paramètre vectoriel  $\Delta^m$  à valeurs dans  $\mathbb{R}^k$  sur lequel on dispose justement de la probabilité "a priori"  $\eta$ . (Souvent  $k = p$  ou un multiple entier de  $p$ ).

5° Nous noterons  $\mathcal{D}$  l'ensemble des fonctions (ou règles) de décision aléatoires

$$\mathcal{D} = \{d / d : (\mathbb{R}^p)^n \rightarrow \mathcal{K}\}$$

où  $\mathcal{K}$  est un ensemble de probabilités sur  $\mathcal{D}$ . On peut identifier  $d$  (voir Wald, A.(1950) chapitre 5) à une application  $d : (\mathbb{R}^p)^n \rightarrow [0, 1]^{m+1}$

$$d = (d_0, d_1, \dots, d_m) \quad \text{avec} \quad 0 \leq d_\nu(x_1, \dots, x_n) \leq 1$$

$\sum_{\nu=0}^m d_\nu(x_1, \dots, x_n) = 1$  ;  $d_\nu(x_1, \dots, x_n)$  représente la probabilité de prendre

la décision  $D_\nu$  sachant  $(x_1, \dots, x_n)$ . Nous noterons:  $\underline{x} = (x_1, \dots, x_n)$  et  $d\underline{x} = dx_1 \dots dx_n$ . De plus pour  $\nu = 0, 1, \dots, m$  nous supposerons que  $d_\nu(\underline{x})$  est mesurable par rapport à la mesure de Lebesgue sur  $(\mathbb{R}^{n \cdot p})$  ; ceci pour pouvoir définir le risque de Bayes généralisé que nous appellerons risque double de Bayes.

6° Nous appellerons  $L$  la fonction de coût (ou fonction de perte).  $L$  est une application de  $\mathcal{K} \times \mathcal{D} \rightarrow \mathbb{R}^+$ .  $L(D_\nu/H_\xi)$  est le coût de la décision  $D_\nu$  lorsque l'état  $H_\xi$  est réalisé.

### §1 EXTENSION DU RISQUE DE BAYES

Nous allons maintenant définir l'analogue, avec nos éléments, du risque de Bayes tel que l'envisageait Wald, A.(1950).

N.B. Nous emploierons indifféremment règle double de Bayes ou règle de Bayes généralisée.

II.1. Définition.

On appelle risque double de Bayes associé à la règle de décision  $d$ , à la probabilité "a priori"  $b_{\theta \eta}$  sur  $\mathcal{E}$ , et à la fonction de coût  $L$  le réel positif que l'on notera  $\bar{R}(d, b_{\theta \eta}, L)$  défini par :

$$\begin{aligned} \bar{R}(d, b_{\theta \eta}, L) &= \int_{\mathcal{E}} \sum_{v=0}^m L(D_v/H_{\xi}) \int_{(\mathbb{R}^p)^n} d_v(\underline{x}) f_{\xi}(\underline{x}, \Delta m) d\underline{x} db_{\theta \eta} \\ &= \sum_{\xi=0}^m b_{\xi} \int_{\mathbb{R}^k} \left[ \sum_{v=0}^m L(D_v/H_{\xi}) \int_{(\mathbb{R}^p)^n} d_v(\underline{x}) f_{\xi}(\underline{x}, \Delta m) d\underline{x} \right] d\eta(\Delta m) \end{aligned}$$

II.2. Définition.

On appellera règle double de Bayes relative à la probabilité a priori  $b_{\theta \eta}$  sur  $\mathcal{E}$ , une règle de décision  $d = (d_0, d_1, \dots, d_m)$  pour laquelle  $\bar{R}(d, b_{\theta \eta}, L)$  est minimal. i.e.  $\forall \bar{d} \in \mathbb{D}, \bar{R}(d, b_{\theta \eta}, L) \leq \bar{R}(\bar{d}, b_{\theta \eta}, L)$

II.3. Définition.

Une règle de décision  $d$  est dite règle double de Bayes s'il existe une probabilité a priori  $b_{\theta \eta}$  sur  $\mathcal{E}$  telle que  $d$  soit une règle double de Bayes relative à  $b_{\theta \eta}$

§2 CARACTERISATION DES REGLES DOUBLES DE BAYES DANS LE CAS D'UNE FONCTION DE COUT  $L_0$  SIMPLE

Nous allons envisager les définitions précédentes avec la fonction de coût suivante :

$$L_0(D_v/H_{\xi}) = \begin{cases} 0 & \text{si } v=\xi \\ 1 & \text{si } v \neq \xi \end{cases}$$

Alors on a :

$$(II.4) \quad \bar{R}(d, b_{\theta \eta}, L_0) = \sum_{\xi=0}^m b_{\xi} \int_{\mathbb{R}^k} \left[ \sum_{\substack{v=0 \\ v \neq \xi}}^m \int_{(\mathbb{R}^p)^n} d_v(\underline{x}) f_{\xi}(\underline{x}, \Delta m) d\underline{x} \right] d\eta(\Delta m)$$

Nous nous proposons de rechercher une propriété des règles de Bayes qui puisse les définir à une équivalence près.

Le risque de Bayes  $\bar{R}$  peut encore s'écrire (théorème de Fubini et linéarité de l'intégrale)

$$\bar{R}(d, b_{0\eta}, L_0) = \int_{\mathbb{R}^{p \cdot n}} \sum_{v=0}^m d_v(\underline{x}) \left[ \sum_{\substack{\xi=0 \\ \xi \neq v}}^m b_\xi \int_{\mathbb{R}^k} f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) \right] d\underline{x}$$

Le risque double de Bayes  $\bar{R}$  sera minimal si et seulement si la quantité

$$\sum_{v=0}^m d_v(\underline{x}) \left[ \sum_{\substack{\xi=0 \\ \xi \neq v}}^m b_\xi \int_{\mathbb{R}^k} f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) \right]$$

est minimal. C'est l'objet du résultat suivant :

posons pour tout  $v = 0, 1, \dots, m$

$$g_v(\underline{x}) = \sum_{\substack{\xi=0 \\ \xi \neq v}}^m b_\xi \int_{\mathbb{R}^k} f_\xi(\underline{x}, \Delta m) d\eta(\Delta m)$$

### II.5. Théorème.

Une condition nécessaire et suffisante pour que  $d = (d_0, d_1, \dots, d_m)$  soit une règle double de Bayes relative à la probabilité a priori  $b_{0\eta}$  sur  $\mathcal{E}$  est :  
 $\forall \underline{x} \in \mathbb{R}^{p \cdot n}$  tel que  $g_v(\underline{x}) > \text{Min} [g_0(\underline{x}), g_1(\underline{x}), \dots, g_m(\underline{x})]$ , on a  $d_v(\underline{x}) = 0$ , sauf peut-être sur un sous-ensemble  $A_0$  de  $\mathbb{R}^{p \cdot n}$  de \* mesure nulle.

N.B.  $\text{Min} [\dots]$  désigne le minimum des  $m+1$  quantités réelles entre crochets.

Par un sous-ensemble  $A_0$  de  $\mathbb{R}^{p \cdot n}$  de \* mesure nulle nous signifions un sous-ensemble  $A_0$  de  $\mathbb{R}^{p \cdot n}$  tel que

$$\int_{\mathcal{E}} \int_{\mathbb{R}^{p \cdot n}} f_v(\underline{x}, \Delta m) d\underline{x} db_{0\eta} = 0$$

Preuve :

Soit  $d$  une règle de décision. Pour tout indice  $v = \{0, 1, \dots, m\}$

considérons le sous-ensemble de  $\mathbb{R}^{p \cdot n}$  :  $A_v(d) = \{\underline{x} \in \mathbb{R}^{p \cdot n} / g_v(\underline{x}) > \text{Min}[g_0(\underline{x}), \dots, g_m(\underline{x})]$

et  $d_v(\underline{x}) \neq 0\}$  supposons qu'il existe un indice  $v_0 \in \{0, 1, \dots, m\}$  tel que

$A_{v_0}(d)$  soit de \* mesure non nulle. Alors considérons la fonction

$\bar{d}^0 : \mathbb{R}^{p \cdot n} \rightarrow [0, 1]^{m+1}$  définie comme suit



II.6. Remarque.

L'éventualité où le  $\text{Min} [g_0(x), g_1(x), \dots, g_m(x)]$  est atteint par plusieurs valeurs à la fois peut se produire. Désignons par  $M$  l'ensemble des  $\underline{x} \in R^{p \cdot n}$  où cela arrive :

$M = \{\underline{x} \in R^{p \cdot n} / \text{Min}[g_0(\underline{x}), g_1(\underline{x}), \dots, g_m(\underline{x})]$  est atteint par plusieurs valeurs}}

$M$  est inclus dans la réunion d'ensembles  $M(v_1, v_2)$  où le minimum est atteint par au moins les deux valeurs  $g_{v_1}(\underline{x})$  et  $g_{v_2}(\underline{x})$  :

$$\begin{aligned} M(v_1, v_2) &= \{\underline{x} \in R^{p \cdot n} / \text{Min}[g_0(\underline{x}), g_1(\underline{x}), \dots, g_m(\underline{x})] \\ &= g_{v_1}(\underline{x}) = g_{v_2}(\underline{x}) \text{ avec } v_1 \neq v_2 \quad v_1 \text{ et } v_2 \text{ éléments de} \\ &\{0, 1, \dots, m\} \cdot \} \end{aligned}$$

Or  $M(v_1, v_2)$  est de \*probabilité nulle.  $M$  est donc inclus dans une réunion finie d'ensembles de \*probabilité nulle, donc  $M$  est de \*probabilité nulle.

Les valeurs des fonctions de décision sur cet ensemble ne changent pas la valeur du risque double de Bayes relatif à la probabilité a priori  $b_{0\eta}$  et peuvent donc être fixées arbitrairement.

Autre forme du théorème II.5.

En explicitant la valeur de  $g_v(\underline{x})$  et en utilisant la remarque ci-dessus nous allons pouvoir donner une équivalence qui nous permettra de caractériser une règle double de Bayes presque partout. Le théorème II.5. s'écrit :  $d$  est une règle double de Bayes relative à la probabilité a priori  $b_{0\eta}$  si et seulement si :

$$(II.7) \quad \left\{ \begin{array}{l} \forall \underline{x} \in R^{p \cdot n}, \text{ sauf peut-être sur un sous-ensemble } A_0 \subset R^{p \cdot n} \text{ de} \\ \text{*probabilité nulle, et pour tout indice } v \text{ tel que } g(\underline{x}) \\ g_v(\underline{x}) > \text{Min}[g_0(\underline{x}), g_1(\underline{x}), \dots, g_m(\underline{x})] \text{ on a : } d_v(\underline{x}) = 0 \end{array} \right.$$

i.e.  $\forall \underline{x} \in R^{p \cdot n} - A_0$ , et pour tout indice  $v$  tels que

$$\sum_{\xi \neq v} b_\xi \int f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) > \text{Min} \left[ \sum_{\xi \neq 0} b_\xi \int f_\xi(\underline{x}, \Delta m) d\eta(\Delta m), \dots, \sum_{\xi \neq m} b_\xi \int f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) \right]$$

on a :  $d_v(\underline{x}) = 0$

$$\text{or } \sum_{\xi \neq \nu} b_{\xi} \int f_{\xi}(\underline{x}, \Delta m) d\eta(\Delta m) = \sum_{\xi=0}^m b_{\xi} \int f_{\xi}(\underline{x}, \Delta m) d\eta(\Delta m) - b_{\nu} \int f_{\nu}(\underline{x}, \Delta m) d\eta(\Delta m)$$

la première quantité après le signe = est indépendante de  $\nu$ .

On peut donc remplacer la condition (II,7) ci-dessus par la condition (II,8) ci-dessous :

$$(II.8) \left\{ \begin{array}{l} \forall \underline{x} \in \mathbb{R}^{p \cdot n}, \text{ sauf peut-être sur un sous-ensemble de } \mathbb{R}^{p \cdot n} \text{ de } b_{0\eta} \text{ proba-} \\ \text{bilité nulle, et } \forall \nu \in \{0, 1, \dots, m\} \text{ tels que} \\ b_{\nu} \int f_{\nu}(\underline{x}, \Delta m) d\eta(\Delta m) < \text{Max} \left[ b_0 \int f_0(\underline{x}, \Delta m) d\eta(\Delta m), \dots, b_m \int f_m(\underline{x}, \Delta m) d\eta(\Delta m) \right] \\ \text{on a } d_{\nu}(\underline{x}) = 0 \end{array} \right.$$

Or l'ensemble :

$$\left\{ \underline{x} \in \mathbb{R}^{p \cdot n} / \text{Max} \left[ b_0 \int f_0(\underline{x}, \Delta m) d\eta(\Delta m), \dots, b_m \int f_m(\underline{x}, \Delta m) d\eta(\Delta m) \right] \right. \\ \left. = b_{\nu_1} \int f_{\nu_1}(\underline{x}, \Delta m) d\eta(\Delta m) = b_{\nu_2} \int f_{\nu_2}(\underline{x}, \Delta m) d\eta(\Delta m) \right. \\ \left. \nu_1 \text{ et } \nu_2 \in \{0, 1, \dots, m\} \right\} = M(\nu_1, \nu_2)$$

Donc l'ensemble des  $\underline{x}$  pour lesquels le maximum est atteint par plusieurs valeurs à la fois est inclus dans  $M$  de \* probabilité nulle.

On peut donc caractériser les règles doubles de Bayes de la façon suivante :

$$II.9 \left\{ \begin{array}{l} \forall \nu \in \{0, 1, \dots, m\} \\ d_{\nu}(\underline{x}) = 1 \text{ si} \\ b_{\nu} \int_{\mathbb{R}^k} f_{\nu}(\underline{x}, \Delta m) d\eta(\Delta m) > \text{Max}_{\substack{\xi \in \{0, 1, \dots, m\} \\ \xi \neq \nu}} \left[ b_{\xi} \int_{\mathbb{R}^k} f_{\xi}(\underline{x}, \Delta m) d\eta(\Delta m) \right] \\ \text{pour tout } x \in \mathbb{R}^{p \cdot n} \\ \text{sauf peut-être sur un sous-ensemble } A_0 \text{ de } * \text{ probabilité nulle;} \\ d \text{ ainsi est définie presque partout.} \end{array} \right.$$

On obtient donc une règle simple et unique (à une équivalence près). La valeur de  $d$  sur  $A_0$  peut-être fixée de façon arbitraire. On en conclut la remarque ci-dessous :

II.10. Remarque.

Rechercher la règle double de Bayes relative à la probabilité a priori  $b_{0\eta}$ , revient donc à déterminer la partition unique à un ensemble de mesure nulle près de l'ensemble des réalisations  $\mathbb{R}^{p \cdot n}$ .

$$\begin{aligned} & E_0, E_1, \dots, E_m, F \\ \text{tels que} \quad E_\nu &= \{ \underline{x} \in \mathbb{R}^{p \cdot n} / b_\nu \int f_\nu(\underline{x}, \Delta m) d\eta(\Delta m) > b_\xi \int f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) \\ & \quad \xi \in \{0, 1, \dots, m\} \quad \xi \neq \nu \} \\ &= \bigcup_{\substack{\xi=0 \\ \xi \neq \nu}}^m \{ \underline{x} \in \mathbb{R}^{p \cdot n} / b_\nu \int f_\nu(\underline{x}, \Delta m) d\eta(\Delta m) > b_\xi \int f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) \} \\ F &= \bigcup_{\nu=0}^m \bar{E}_\nu - E_\nu \quad \text{où } \bar{E}_\nu \text{ désigne la fermeture de } E_\nu \text{ pour } \nu=0 \end{aligned}$$

$\nu = 0, 1, \dots, m$ .  $F$  est la réunion des points frontière ;  $F$  est de probabilité nulle.

$$\begin{aligned} \text{Alors } P_{(H_\xi, \Delta m)}(D_\nu) &= P[\underline{x} \in E_\nu / (H_\xi, \Delta m)] \\ &= \int d(\underline{x}) f_\xi(\underline{x}, \Delta m) dx \end{aligned}$$

N.B. ne pas confondre  $A$  et  $F$ .

§3 CARACTERISATION DANS UN CAS PARTICULIER

Nous allons traduire le résultat II.9 ci-dessus dans le cas où

- 1° la densité  $f_0$  ne dépend que de  $\underline{x}$  et pas de  $\Delta m$
- 2° la probabilité a priori  $b$  sur  $\mathcal{H}$  nous est donnée par

$$\begin{aligned} b[H_0] &= b \\ b[H_\nu] &= b_\nu = \frac{1-b}{m} \quad \text{pour tout } \nu \neq 0 \end{aligned}$$

c'est-à-dire que nous considérons les hypothèses  $H_\nu$  pour  $\nu = 1, \dots, m$  comme équiprobables.

$$\text{Alors } \int_{\mathbb{R}^k} f_0(\underline{x}, \Delta m) d\eta(\Delta m) = f_0(\underline{x}) \int_{\mathbb{R}^k} d\eta(\Delta m) = f_0(\underline{x})$$

De plus, en posant  $A = \frac{mb_0}{1-b_0}$  on a  $b_0 \in ]0, 1[ \Rightarrow A \in ]0, +\infty[$

En multipliant l' inégalité II.9 par la constante positive  $\frac{m}{1-b_0}$  on obtient la caractérisation suivante des règles de Bayes généralisées relatives à la probabilité  $b_{\otimes \eta}$

$$(II.11.) \quad \text{Caractérisation} \quad \left\{ \begin{array}{l} d_0(\underline{x}) = 1 \text{ si } Af_0(\underline{x}) > \int f_v(\underline{x}, \Delta m) d\eta(\Delta m) \text{ pour tout } v = 1 \dots m \\ d_v(\underline{x}) = 1 \text{ si } \left\{ \begin{array}{l} Af_0(\underline{x}) < \int f_v(\underline{x}, \Delta m) d\eta(\Delta m) \text{ et} \\ \int f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) < \int f_v(\underline{x}, \Delta m) d\eta(\Delta m) \\ \text{pour } \xi \in \{1, \dots, m\} \quad \xi \neq v \end{array} \right. \end{array} \right.$$

pour les autres points  $\underline{x}$  de  $\mathbb{R}^{p \cdot n}$   $d$  peut-être fixée arbitrairement.

#### §4 AUTRE PROPRIETE CARACTERISTIQUE

Nous donnons ici, toujours dans le cas de la fonction de coût  $L_0$ , une nouvelle propriété des règles doubles de Bayes.

Posons :

$$I(d, b_{\otimes \eta}) = \int_{\mathbb{R}^{p \cdot n}} \sum_{v=0}^m b_v d_v \int f_v(\underline{x}, \Delta m) d\eta(\Delta m) d\underline{x}$$

#### II.12. Proposition.

L'intégrale  $I(d, b_{\otimes \eta})$  est maximale si et seulement si  $d$  est une règle double de Bayes.

Preuve : considérons l'intégrale suivante :

$$\begin{aligned} J(d, b_{\otimes \eta}) &= \int_{\mathbb{R}^{p \cdot n}} \sum_{\xi=0}^m b_\xi \left( \sum_{v=0}^m d_v \int f_\xi(\underline{x}, m) d\eta(m) \right) d\underline{x} \\ &= \int_{\mathbb{R}^{p \cdot n}} \sum_{\xi=0}^m b_\xi \left( \sum_{\substack{v=0 \\ v \neq \xi}}^m d_v \int f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) \right) d\underline{x} \end{aligned}$$

$$+ \int_{\mathbb{R}^{p \cdot n}} \sum_{v=0}^m b_v d_v \left[ f_v(\underline{x}, \Delta m) d\eta(\Delta m) \right] d\underline{x}$$

$$= \bar{R}(d, b_{0\eta}, L_0) + I(d, b_{0\eta})$$

Mais  $J(d, b_{0\eta}) = \int_{\mathbb{R}^{p \cdot n}} \left( \sum_{v=0}^m d_v \right) \sum_{\xi=0}^m b_\xi \left[ f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) \right] d\underline{x}$

$$= \int_{\mathbb{R}^{p \cdot n}} \sum_{\xi=0}^m b_\xi \left[ f_\xi(\underline{x}, \Delta m) d\eta(\Delta m) \right] d\underline{x}$$

$J$  est donc indépendante de la fonction de décision  $d$  et si  $d$  et  $\bar{d}$  désignent deux fonctions de décision relatives à la probabilité a priori  $b_{0\eta}$  sur  $\mathcal{E}$  et la fonction de coût  $L_0$ , on a

$$(II.13) \quad J(d, b_{0\eta}) = J(\bar{d}, b_{0\eta})$$

Or  $d$  est une règle double de Bayes relative à la probabilité a priori  $b_{0\eta}$  si et seulement si  $\forall \bar{d} \in \mathcal{D}, \bar{R}(d, b_{0\eta}, L_0) \leq \bar{R}(\bar{d}, b_{0\eta}, L_0)$ . En combinant ce résultat avec l'égalité ((II.13)) ci-dessus, on conclut que  $d$  est une règle de Bayes relative à la probabilité a priori  $b_{0\eta}$  si et seulement si

$$(II.14.) \quad \forall \bar{d} \in \mathcal{D}, I(d, b_{0\eta}) \gg I(\bar{d}, b_{0\eta})$$

### II.15. Remarque.

En d'autres termes, rechercher les règles doubles de Bayes relatives à la probabilité a priori  $b_{0\eta}$  et la fonction de coût  $L_0$  est équivalent à vouloir rendre l'intégrale  $I(d)$  maximale.

### II.16. Corollaire.

Lorsque  $f_0$  ne dépend pas de  $\Delta m$  et que  $b$  est la probabilité a priori sur  $\mathcal{E}$  définie au §3 l'intégrale

$$I(d) = \int A d_0(\underline{x}) f_0(\underline{x}) + \sum_{v=1}^m d_v(\underline{x}) \left[ f_v(\underline{x}, \Delta m) d\eta(\Delta m) \right] d\underline{x}$$

où  $A$  est une constante positive, est maximale si et seulement si  $d$  est une règle double de Bayes relative à la probabilité a priori  $b_{0\eta}$ .

(Preuve immédiate : il s'agit seulement de traduire les hypothèses)

## §5 PROPRIÉTÉ DES RÈGLES DOUBLES DE BAYES SYMÉTRIQUES R.P.D.

Nous nous intéressons maintenant à une classe de fonctions de décision particulières et démontrons une propriété qui nous servira à plusieurs reprises par la suite (chapitre III et IV).

Nous maintenons les hypothèses du corollaire II.16 : i.e la densité  $f_0$  ne dépend pas de  $\Delta m$  et  $b$ . Nous nous situons donc sur le sous-ensemble de  $\mathcal{C}$  suivant :

$$(H_0 \times O) \cup \{H_1, \dots, H_m\} \times \mathbb{R}^k.$$

Nous nous limitons aux fonctions de décision qui ont la même probabilité de décision pour l'hypothèse  $H_0$  :

$$P_{H_0}(D_0) = 1 - a \quad \text{où } a \in ]0, 1[ \quad a \text{ étant fixé indépendamment de } d.$$

Nous donnons de plus deux définitions :

### II.17. Définition.

On appelle règle de décision symétrique\* une règle de décision pour laquelle la probabilité de décision  $P_{(H_v, \Delta m)}(D_v)$  de prendre la décision correcte  $D_v$  lorsque l'état  $(H_v, \Delta m)$  est réalisé ne dépend pas de  $v = 1 \dots m$ .

N.B. Lorsqu'il y aura risque de confusion nous noterons également  $P(D_v/d, (H_v, \Delta m))$  la probabilité  $P_{H_v, \Delta m}(D_v)$  pour la règle de décision  $d$ .

### II.18. Définition

Une règle  $d$  est dite règle double de Bayes symétrique\* si parmi les règles symétriques\*  $d$  est une règle double de BAYES

Enfin nous formulons la condition supplémentaire suivante :

\* relativement aux probabilités de décision

II.19. Hypothèse.

Soit  $\Lambda^{-1}$  la matrice associée à une forme quadratique définie positive ; alors pour une règle de décision donnée  $d \in a$  la classe à laquelle nous nous restreignons,  $P(D_v/d, (H_v, \Delta m))$  ne dépend que pour  $||\Delta m||_{\Lambda^{-1}}$  où

$$||\Delta m||_{\Lambda^{-1}}^2 = \Delta m' \Lambda^{-1} \Delta m$$

Pour  $c$  réel positif fixé nous définissons la mesure  $\eta_c(\Delta m)$  comme une mesure, absolument continue par rapport à la mesure de Lebesgue, concentrée sur l'ellipsoïde :  $\Delta m' \Lambda^{-1} \Delta m = c$ . Nous pouvons alors énoncer le résultat suivant :

II.20. Théorème.

Parmi les règles de décision symétriques\* vérifiant les hypothèses du corollaire II.16. et l'hypothèse II.19. la règle double de Bayes symétrique\*  $d$ , maximise pour toute constante  $A$  et toute mesure  $\eta_c$  définie plus haut la probabilité de décision  $P_{(H_v, \Delta m)}(D_v)$  uniformément en  $v = 1, \dots, m$ .

(Preuve :

Soit  $A$  une constante positive et  $\eta_c$  une mesure telle que

$$\eta_c \{ \Delta m' \Lambda^{-1} \Delta m \} = 1. \text{ Supposons qu'il existe une règle de décision}$$

$\bar{d}$  vérifiant les hypothèses du théorème II.20. un indice  $v_0$  et un vecteur  $\Delta m$  de  $(\mathbb{R}^k)^*$  de norme  $||\Delta m||_{\Lambda^{-1}}^2 = c$  tel que

$$P(D_{v_0}/d, (H_{v_0}, \Delta m)) < P(D_{v_0}/\bar{d}, (H_{v_0}, \Delta m))$$

Alors d'après la symétrie, cette inégalité est vérifiée pour tout  $v = 1 \dots m$  et d'après l'hypothèse II.19. pour tout  $\Delta m$  de  $\Lambda^{-1}$ -norme :  $c^{1/2}$ , soit

$$\forall v = 1 \dots m, \forall \Delta m \in \{ \Delta m / ||\Delta m||_{\Lambda^{-1}}^2 = c \}$$

$$\int_{\mathbb{R}^{p \cdot n}} d_v(x) f_v(x, \Delta m) dx < \int_{\mathbb{R}^{p \cdot n}} \bar{d}_v(x) f_v(x, \Delta m) dx$$

d'autre part, par hypothèse

$$\int_{\mathbb{R}^{p \cdot n}} d_0(x) f_0(x) dx = \int_{\mathbb{R}^{p \cdot n}} \bar{d}_0(x) f_0(x) dx = 1 - a$$

\* relativement aux probabilités de décision

alors

$$\int A \int_{\mathbb{R}^{p \cdot n}} d_0(\underline{x}) f_0(\underline{x}) d\underline{x} + \sum_{v=1}^m \int_{\mathbb{R}^{p \cdot n}} d_v(\underline{x}) f_v(\underline{x}, \Delta m) d\underline{x} d\eta_c(\Delta m) < \int A \int_{\mathbb{R}^{p \cdot n}} \bar{d}_0(\underline{x}) f_0(\underline{x}) d\underline{x} + \sum_{v=1}^m \int_{\mathbb{R}^{p \cdot n}} \bar{d}_v(\underline{x}) f_v(\underline{x}, \Delta m) d\underline{x} d\eta_c(\Delta m)$$

d'où en appliquant le théorème de Fubini :

$$\int_{\mathbb{R}^{p \cdot n}} A d_0(\underline{x}) f_0(\underline{x}) + \sum_{v=1}^m d_v(\underline{x}) \int f_v(\underline{x}, \Delta m) d\eta(\Delta m) d\underline{x} < \int_{\mathbb{R}^{p \cdot n}} A \bar{d}_0(\underline{x}) f_0(\underline{x}) + \sum_{v=1}^m \bar{d}_v(\underline{x}) \int f_v(\underline{x}, \Delta m) d\eta(\Delta m) d\underline{x}$$

ce qui contredit la proposition II.12.)

### §6 QUELQUES RESULTATS SUR L'INVARIANCE

Les tests que nous désirons construire doivent être symétriques en  $x_1, \dots, x_n$ . L'expression mathématique de la symétrie est l'invariance sous un certain groupe  $G$  de transformations. Nous rappelons ici certaines définitions de l'invariance en nous inspirant de Lehmann, E.L.(1959). Puis nous donnons certains résultats concernant les règles de décision invariantes qui sont des généralisations de certains résultats de Kudo, A.(1956 b).

En particulier, nous caractérisons les règles doubles de Bayes lorsqu'on se limite aux règles de décision invariantes.

Soit  $\mathbb{R}^{p \cdot n}$  l'espace des échantillons. Soit  $G$  un groupe de transformations de  $\mathbb{R}^{p \cdot n}$  sur lui-même.

Nous définissons sur  $\mathbb{R}^{p \cdot n}$  la relation d'équivalence suivante.

Soient  $\underline{x}^1, \underline{x}^2 \in \mathbb{R}^{p \cdot n}$

$$(II.21.) \quad \underline{x}^1 \sim \underline{x}^2 \pmod{G} \iff \exists g \in G \text{ tel que } \underline{x}^2 = g(\underline{x}^1)$$

II.22. Définition.

Les classes d'équivalence de la relation d'équivalence II.21. sont appelées des orbites : i.e.

$$\underline{x}^1 = \{ \underline{x} \in \mathbb{R}^{p.n} / \exists g \in G \text{ tel que } \underline{x} = g(\underline{x}^1) \}$$

Les orbites constituent une partition de  $\mathbb{R}^{p.n}$

II.23. Définition.

Une fonction T définie sur  $\mathbb{R}^{p.n}$  est invariante sous le groupe G, si et seulement si elle est constante sur les orbites.

II.24. Définition.

Nous dirons qu'une fonction T définie sur  $\mathbb{R}^{p.n}$  est un invariant maximal si et seulement si T est invariante sous G et si pour tout élément  $\underline{x}^1$  et  $\underline{x}^2$  de  $\mathbb{R}^{p.n}$

$$T(\underline{x}^1) = T(\underline{x}^2) \implies \exists g \in G \text{ tel que } \underline{x}^2 = g(\underline{x}^1)$$

C'est-à-dire qu'une fonction est un invariant maximal si et seulement si elle est constante sur les orbites et prend une valeur distincte sur chacune d'elles. Les invariants maximaux sont équivalents en ce sens que les ensembles où ils sont constants sont les mêmes.

Rappelons également le théorème suivant de Lehman, L.E. (1959).

II.25. Théorème.

Soit G un groupe de transformations et soient D et E deux sous-groupes de G qui l'engendrent. Supposons que  $y = s(\underline{x})$  est un invariant maximal par rapport à D et que  $\forall e \in E ; \forall \underline{x}^1, \underline{x}^2 \in \mathbb{R}^{p.n}$

$$s(\underline{x}^1) = s(\underline{x}^2) \implies s(e\underline{x}^1) = s(e\underline{x}^2)$$

Si  $z = t(y)$  est un invariant maximal sous  $E^*$  groupe des transformations  $e^*$  définies par  $e^* y = s(e\underline{x})$  quand  $y = s(\underline{x})$  alors  $z = t(s(\underline{x}))$  est un invariant maximal sous le groupe de transformations G.

Nous supposons dans ce qui suit que la variable aléatoire  $X$  associée aux observations  $x_1, \dots, x_n$  est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^p$ .

$\mathcal{V}$  désignera une fonction de décision.

Nous allons nous restreindre à la classe de transformations  $g$  telles que

$$\forall \underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^{p \cdot n}$$
$$g(\underline{x}) = (Bx_1 + c, \dots, Bx_n + c)$$

où  $B$  est un élément du groupe linéaire  $GL(p, \mathbb{R})$  et  $c$  un vecteur constant de  $\mathbb{R}^p$ .

### II.26. Définition.

Nous appellerons  $G_1, G_2, G_3$  les groupes de transformations définies ci-dessus vérifiant respectivement

1°  $B = I_p$

2°  $c = 0$

3°  $B$  et  $c$  quelconques

### II.27. Remarque.

$G_1$  est le groupe des translations de chacun des  $x_\mu$  pour  $\mu = 1, \dots, n$  d'un vecteur constant de  $\mathbb{R}^p$ .

$G_2$  est le groupe des multiplications de chacun des  $x_\mu$  par une matrice inversible.

$G_3$  est le groupe engendré par  $G_1$  et  $G_2$ .

### II.28. Proposition.

La fonction  $s$  définie sur  $\mathbb{R}^{p \cdot n}$  par  $\forall \underline{x} \in \mathbb{R}^{p \cdot n}$

$$s(\underline{x}) = (x_2 - x_1, \dots, x_n - x_1)$$

est un invariant maximal sous  $G_1$ , sur le sous-ensemble de  $\mathbb{R}^{p \cdot n}$  de mesure 1 où  $x_1$  est fini.

(Preuve : tout d'abord, montrons que  $s$  est invariante : en effet soit  $c$  un vecteur quelconque de  $\mathbb{R}^p$  associé à  $g_1 \in G_1$ . Alors

$$\begin{aligned} \forall \underline{x} \in \mathbb{R}^{p \cdot n}, s(g_1 \underline{x}) &= s(x_1 + c, \dots, x_n + c) \\ &= (x_2 - x_1, \dots, x_n - x_1) \\ &= s(\underline{x}) \end{aligned}$$

D'autre part, montrons que  $s$  prend une valeur distincte sur chaque orbite. Pour cela, supposons qu'il existe  $\underline{x}^1$  et  $\underline{x}^2 \in \mathbb{R}^{p \cdot n}$  tels que

$$s(\underline{x}^1) = s(\underline{x}^2)$$

i.e.  $(x_2^1 - x_1^1, \dots, x_n^1 - x_1^1) = (x_2^2 - x_1^2, \dots, x_n^2 - x_1^2)$

i.e.  $\forall \mu = 2, \dots, n$

$$x_\mu^1 - x_1^1 = x_\mu^2 - x_1^2$$

Posons  $x_1^2 - x_1^1 = c \in \mathbb{R}^p$  et soit  $g_1$  l'élément de  $G_1$  associé à  $c$ , alors

$$\forall \mu = 1, \dots, n \quad x_\mu^2 = x_\mu^1 + c$$

soit  $\underline{x}^2 = g_1(\underline{x}^1)$  donc  $\underline{x}^2 \sim \underline{x}^1$ .)

### II.29. Remarque.

On aurait pu également montrer que la fonction :

$$(x_1, \dots, x_n) \rightsquigarrow (x_1 - x_n, \dots, x_{n-1} - x_n)$$

est un invariant maximal sous  $G_1$

### II.30. Proposition.

La fonction  $t$  définie par

$$t(\underline{x}) = B^{-1}x_1, \dots, B^{-1}x_{n-p} \quad \text{où } B = (x_{n-p+1}, \dots, x_n),$$

est un invariant maximal par rapport à  $G_2$ , sur le sous-ensemble de  $\mathbb{R}^{p \cdot n}$  où  $B$  est inversible

### II.31. Remarque.

Les observations  $x_1, \dots, x_n$  sont stochastiquement indépendantes. Donc avec une probabilité 1,  $x_{n-p+1}, \dots, x_n$  sont  $p$  vecteurs de  $\mathbb{R}^p$  linéairement indépendants. Donc  $B$  est inversible presque partout.

(Preuve de la proposition II.30.

Tout d'abord montrons que  $t$  est invariante sous  $G_2$ . Soient  $A \in GL(p, \mathbb{R})$   
 $\underline{x} \in \mathbb{R}^{p \cdot n}$ .

Nous noterons  $g_2$  la transformation de  $\mathbb{R}^{p \cdot n}$  associée à  $A$ .

Alors

$$t(A\underline{x}) = B^{-1}Ax_1, \dots, B^{-1}Ax_{n-p}$$

où  $B$  est la matrice formée des  $p$  vecteurs colonnes  $Ax_{n-p+1}, \dots, Ax_n$

soit  $B = A(x_{n-p+1}, \dots, x_n)$ . Alors

$$B^{-1} = (x_{n-p+1}, \dots, x_n)^{-1} A^{-1} \text{ d'où}$$

$$\begin{aligned} t(A\underline{x}) &= (x_{n-p+1}, \dots, x_n)^{-1} x_1, \dots, (x_{n-p+1}, \dots, x_n)^{-1} x_{n-p} \\ &= t(\underline{x}) \end{aligned}$$

Montrons de plus que  $t$  est distincte sur chaque orbite pour cela soient  
 $\underline{x}^1$  et  $\underline{x}^2 \in \mathbb{R}^{p \cdot n}$  tels que

$$t(\underline{x}^1) = t(\underline{x}^2)$$

$$\underline{x}^1 = (x_1^1, \dots, x_n^1) \quad \underline{x}^2 = (x_1^2, \dots, x_n^2)$$

Nous noterons  $A_1$  la matrice  $\in GL(p, \mathbb{R})$  p.s.

$$A_1 = (x_{n-p+1}^1, \dots, x_n^1)$$

et  $A_2$  la matrice  $\in GL(p, \mathbb{R})$  p.s.

$$A_2 = (x_{n-p+1}^2, \dots, x_n^2)$$

Soit  $\mathcal{G} = A_2 \times A_1^{-1}$  où  $\times$  désigne le produit matriciel ordinaire.

Alors  $\mathcal{G} \in GL(p, \mathbb{R})$  p.s.

Par hypothèse  $\forall \mu = 1, \dots, n-p$

$$A_1^{-1}(x_\mu^1) = A_2^{-1}x_\mu^2$$

Alors  $\forall \mu = 1 \dots n$

$$x_\mu^2 = \mathcal{G} x_\mu^1$$

soit en notant  $g_2$  la transformation associée à  $\mathcal{G}$

$$\underline{x}^2 = g_2(\underline{x}^1)$$

II.32. Proposition.

La fonction  $t \circ s$  définie sur  $\mathbb{R}^{p \cdot n}$  par

$$t \circ s(\underline{x}) = B^{-1}(x_2 - x_1), \dots, B^{-1}(x_{n-p} - x_1) \text{ où } B \text{ est la matrice } p \times p$$

formée des  $p$  vecteurs-colonne  $x_{n-p+1} - x_1, \dots, x_n - x_1$  est un invariant maximal sous le groupe de transformations  $G_3$ , sur le sous ensemble de  $\mathbb{R}^{p \cdot n}$  où  $B$  est inversible.

II.33. Remarque.

Les observations  $x_1 \dots x_n$  sont stochastiquement indépendantes. On en conclut que les variables aléatoires  $x_{n-p+1} - x_1, \dots, x_n - x_1$  sont également stochastiquement indépendantes et on applique la remarque II.31.

(Preuve : compte tenu de la remarque II.27. on vérifie les hypothèses du théorème II.25).

II.34. Définition.

Soit  $\psi$  une fonction de décision définie sur  $\mathbb{R}^{p \cdot n}$  l'espace des échantillons. On dira que  $\psi$  est invariante sous le groupe  $G$  si et seulement si

$$\forall g \in G, \psi(gx) = \psi(x)$$

sauf peut-être pour un sous-ensemble  $M \subset \mathbb{R}^{p \cdot n}$  de mesure nulle ( $M$  est fixé pour  $\psi$  donnée, et ne dépend pas de  $g \in G$ )

II.35. Définition.

Nous dirons que  $\psi$  est invariante sous  $G$  au sens large si

$$\forall g \in G \quad \psi(gx) = \psi(x)$$

pour tout  $x \in \mathbb{R}^{p \cdot n}$  sauf peut-être sur un ensemble de  $\bar{x}$  de mesure nulle, cet ensemble étant fonction de  $g \in G$ . et la réunion de ces ensembles n'étant pas forcément de mesure nulle.

II.36. Définition.

Nous appellerons  $G'_1$  (resp.  $G'_2, G'_3$ ) le groupe de transformations  $g'_1$  (resp  $g'_2, g'_3$ ) associé à  $G_1$  (resp.  $G_2, B_3$ ) où cette fois la matrice  $B$  et le vecteur

$C \in \mathbb{R}^p$  sont des fonctions de  $\underline{x} \in \mathbb{R}^{p \cdot n}$  qui vérifient la condition 1° (resp 2°-3°) de la définition II.26 presque partout. L'ensemble de mesure nulle où ces conditions ne sont peut-être pas vérifiées peut dépendre de  $g'_1 \in G'_1$  (resp  $g'_2 \in G'_2, g'_3 \in G'_3$ ). En notant  $M_g$  l'ensemble de mesure nulle associé à la fonction  $g$ , où les conditions ne sont peut-être pas vérifiées et en supposant que  $\bigcup_{g \in G} M_g$  est de mesure nulle, nous avons le résultat suivant :

II.37. Théorème.

Soit  $\varphi$  une fonction de décision sur  $\mathbb{R}^{p \cdot n}$ . Une condition nécessaire et suffisante pour que  $\varphi$  soit invariante sous le groupe de transformations  $G_1$  (resp.  $G_2, G_3$ ) est que  $\varphi$  soit invariante sous le groupe de transformations  $G'_1$  (resp.  $G'_2, G'_3$ )

Nous avons déjà donné la démonstration concernant  $G_1$  et  $G'_1$  (voir Garel, B.B. (1975)). Nous donnons ici la démonstration pour les groupes  $G_3$  et  $G'_3$  étant entendu que la démonstration pour  $G_2$  et  $G'_2$  s'obtient de façon analogue.

(Preuve : de façon évidente  $G_3 \subset G'_3$ . Donc si  $\varphi$  est invariante sous  $G'_3$ ,  $\varphi$  est invariante sous  $G_3$ . Démontrons la réciproque par l'absurde. Supposons  $\varphi$  invariante sous  $G_3$  et non-invariante sous  $G'_3$ . Il existe donc une fonction  $g'_3 \in G'_3$  et un ensemble  $N$  de mesure positive tels que

$$\forall \underline{x} \in N$$

$$\begin{aligned} \varphi(\underline{x}) &= (x_1, \dots, x_n) \neq \varphi(g'_3(\underline{x})) \\ &= \varphi(B_0(\underline{x})x_1 + c_0(\underline{x}), \dots, B_0(\underline{x})x_n + c_0(\underline{x})) \end{aligned}$$

où  $B_0(\underline{x})$  est la matrice fonction de  $\underline{x}$  et  $c_0(\underline{x})$  le vecteur de  $\mathbb{R}^p$  associés à  $g'_3$ ;  $B_0(\underline{x})$  est inversible p.s. et  $c_0(\underline{x})$  fini p.s.

Soit  $M$  l'ensemble de mesure nulle de la définition II.34. associé à  $\varphi$ .

Considérons l'ensemble  $N' = N - (M \cup \bigcup_{g'_3 \in G'_3} N_{g'_3})$ .  $N'$  est de mesure non nulle. D'autre part  $\forall \underline{x}^0 \in N'$

$$\varphi(g'_3(\underline{x}^0)) \neq \varphi(\underline{x}^0)$$

i.e. 
$$\varphi(B_0(\underline{x}^0)x_1^0 + c_0(\underline{x}^0), \dots, B_0(\underline{x}^0)x_n^0 + c_0(\underline{x}^0)) \neq \varphi(\underline{x}^0)$$

Soit alors  $g_3$  la fonction associée à la matrice fixe et inversible  $B_0(\underline{x}^0)$  et le vecteur fixe  $C_0(\underline{x}^0)$  fini .

On a la relation

$$\Psi(g_3(\underline{x}^0)) \neq \Psi(\underline{x}^0)$$

En définitive :

$$\forall \underline{x}^0 \in N' \quad \exists g_3 \in G_3 \quad \text{tel que} \quad \Psi(g_3(\underline{x}^0)) \neq \Psi(\underline{x}^0)$$

ce qui contredit l'hypothèse)

Si l'on enlève l'hypothèse :  $\int_g M_g$  de mesure nulle, nous avons seulement le résultat suivant :

II.38. Proposition.

Soit  $\Psi$  une règle de décision invariante sous  $G_1$  (resp.  $G_2, G_3$ ) . Alors  $\Psi$  est invariante au sens large sous  $G'_1$  (resp.  $G'_2, G'_3$ ).

Nous laissons à la charge du lecteur la démonstration, car elle se calque sur celle du théorème II.37.

Nous allons maintenant préciser ces résultats dans le cadre défini précédemment. L'espace des états de la nature  $\mathcal{E} = \mathcal{H} \times \mathbb{R}^k$  où  $\mathcal{H} = \{H_0, \dots, H_m\}$ . L'hypothèse  $H_\nu$  pour  $\nu = 1 \dots m$  est associée à un paramètre inconnu  $\Delta_m \in \mathbb{R}^k$  la décision  $D_0$  ne dépend pas de  $\Delta_m$  .

L'ensemble  $\{D_0, \dots, D_m\}$  détermine une partition de

$$H_0 \times 0 \cup \{H_1, \dots, H_m\} \times \mathbb{R}^k; \text{ à } D_0 \text{ est associé } H_0 \times 0$$

$$\text{à } D_\nu \text{ est associé } H_\nu \times \mathbb{R}^k \text{ pour } \nu = 1 \dots m .$$

Nous dirons donc que  $D_\nu$  est la décision de retenir l'hypothèse  $H_\nu$  . Nous voulons donc faire un choix entre  $m$  hypothèses  $H_0, \dots, H_m$  auxquelles correspondent les  $m$  décisions  $D_0, \dots, D_m$  . Soit  $d_\nu(\underline{x})$  la probabilité de prendre la décision  $D_\nu$  après avoir observé  $\underline{x}$  pour  $\nu = 0, 1, \dots, m$  .

Alors  $0 \leq d_\nu(\underline{x}) \leq 1$

$$\sum_{\nu=0}^m d_\nu(\underline{x}) = 1$$

Selon la terminologie de Wald, A. (1950) reprise par Kudo, A. (1956 b) nous

appelons règle de décision associée à notre problème la fonction  $d$

$$d(\underline{x}) = (d_0(\underline{x}), \dots, d_m(\underline{x}))$$

II.39. Définition.

La règle de décision  $d$  est invariante sous le groupe de transformation  $G$  si et seulement si pour tout  $v \in \{0, 1, \dots, m\}$  la fonction  $d_v$  est invariante sous  $G$ .

Nous avons obtenu la caractérisation II.11. des règles doubles de Bayes relatives à une probabilité a priori  $b_{0\eta}$  sur  $\mathcal{E}$ .

Nous allons maintenant préciser cette caractérisation en faisant certaines hypothèses d'invariance.

Nous avons vu au §4 du chapitre II proposition II.12. que rechercher les règles double de Bayes est équivalent à maximiser l'intégrale  $I(d, b_{0\eta})$  où

$$I(d, b_{0\eta}) = \int_{\mathbb{R}^{p.n}} A d_0(\underline{x}) f_0(\underline{x}) + \sum_{v=1}^m d_v(\underline{x}) \int f_v(\underline{x}, \Delta m) d\eta(\Delta m) d\underline{x}$$

On obtient alors le résultat suivant :

II.40. Théorème.

Une condition nécessaire et suffisante pour que la règle de décision  $d$  satisfaisant aux hypothèses du corollaire II.16. soit une règle double de Bayes invariante sous le groupe de transformation  $G_1$  (resp.  $G_3$ ) est que

$$\begin{aligned} d_0(\underline{x}) &= 1 \text{ pour tout } \underline{x} \text{ tel que } Ah_0(Y) > \int h_v(Y, \Delta m) d\eta(\Delta m) \text{ pour } v = 1 \dots m \\ d_v(\underline{x}) &= 1 \text{ pour tout } \underline{x} \text{ tel que } \int h_v(Y, \Delta m) d\eta(\Delta m) > Ah_0(Y) \\ &\text{et } \int h_v(Y, \Delta m) d\eta(\Delta m) > \int h_\xi(Y, \Delta m) d\eta(\Delta m) \text{ pour } \xi = 1, \dots, m \\ &\hspace{15em} \xi \neq v \end{aligned}$$

sauf peut-être sur un ensemble de  $\underline{x}$  de mesure nulle.  $h_0(Y)$  (resp.  $(h_v(Y, \Delta m))$ ) est la densité de l'invariant maximal sous le groupe  $G_1$  (resp.  $G_3$ ) sous les hypothèses  $H_0$  et  $H_v$  respectivement.

(Preuve : nous ferons la démonstration dans le cas du groupe  $G_3$  étant entendu que dans le cas du groupe  $G_1$  on obtient la démonstration de façon analogue.

Soit  $d$  une règle de décision satisfaisant aux hypothèses du corollaire II.16. telle que  $d$  soit une règle double de Bayes, relativement à la probabilité a priori  $b_{\eta}$ , invariante sous le groupe  $G_3$ . Considérons l'intégrale

$$I(d, b_{\eta}) = \int [A d_0(\underline{x}) f_0(\underline{x}) + \sum_{v=1}^m d_v(\underline{x}) \int f_v(\underline{x}, \Delta m) d\eta(\Delta m)] d\underline{x}$$

Appliquant à  $d = (d_0, \dots, d_m)$  la proposition II.38. en prenant pour  $g'_3$  la fonction associée à la matrice  $p \times p$

$$B^{-1}(\underline{x}) = (x_{n-p+1}, \dots, x_n)^{-1} \text{ et au vecteur } c(\underline{x}) = -B^{-1}x_1$$

Alors sauf sur un ensemble  $N$  de mesure nulle qui dépend de  $B(\underline{x})$  et  $c(\underline{x})$  on a

$$d(g'_3(\underline{x})) = d(\underline{x}) \text{ p.s.}$$

soit  $\forall v = 0, 1, \dots, m$

$$d_v(g'_3(\underline{x})) = d_v(\underline{x}) \text{ p.s.}$$

avec

$$d_v(\underline{x}) = d_v(x_1, \dots, x_n)$$

$$d_v(g'_3(\underline{x})) = d_v(0, B^{-1}(x_2 - x_1), \dots, B^{-1}(x_{n-p} - x_1), 1, \dots, 1)$$

De façon évidente,  $N$  étant de mesure nulle :

$$I(d, b_{\eta}) = \int_{\mathbb{R}^{p-n-N}} [A d_0(\underline{x}) f_0(\underline{x}) + \sum_{v=1}^m d_v(\underline{x}) \int f_v(\underline{x}, \Delta m) d\eta(\Delta m)] d\underline{x}$$

opérons alors le changement de variable

$$y_2 = B^{-1}(x_2 - x_1)$$

$$\vdots$$

$$y_{n-p} = B^{-1}(x_{n-p} - x_1)$$

$$W = x_1$$

où  $B$  est la matrice définie plus haut inversible sur  $\mathbb{R}^{p-n-N}$

$$I(d, b_{\eta}) = \int_{\mathbb{R}^{p-n-N}} [A d_0(0, y_2, \dots, y_{n-p}, 1, \dots, 1) f_0(W, By_2 + W, \dots, By_{n-p} + W, B + (W) + \sum_{v=1}^m d_v(0, y_2, \dots, y_{n-p}, 1, \dots, 1) \int f_v(W, By_2 + W, \dots, By_{n-p} + W, B + (W), \Delta m) d\eta(\Delta m)] |\det B|^{n-p-1} dY dB dW$$

où  $Y = (y_2, \dots, y_{n-p})$   
 $(W) = (W, W, \dots, W)$  matrice  $p \times p$  formée de  $p$  vecteurs colonnes  $W$ .

(II.42.) Posons  $d_\nu(Y) = d_\nu(0, y_2, \dots, y_{n-p}, 1, \dots, 1)$   
 pour tout  $\nu = 0, 1, \dots, m$  et tout  $Y \in \mathbb{R}^{p \cdot (n-p-1)} \cap N$

alors on a  $I(d, b^{\otimes n}) =$   

$$\int_{\mathbb{R}^{p(n-p-1)}} A d_0(Y) h_0(Y) + \sum_{\nu=1}^m d_\nu(Y) \int h_\nu(Y, \Delta m) d\eta(\Delta m) dY$$

que nous noterons  $I_Y(d, b^{\otimes n})$

où  $h_\nu(Y, m) = \int f_\nu(W, By_2 + W, \dots, By_{n-p} + W, B + (W), \Delta m) |\det B|^{n-p-1} dBdW$   
 est justement la densité de l'invariant maximal  $Y$  sous le groupe  $G_3$ , sous l'hypothèse  $H_\nu$  et  $h_0(Y)$  étant l'analogie sous  $H_0$ .

En appliquant les résultats obtenus en  $\underline{x}$  pour  $I(d, b^{\otimes n})$  on obtient que  $I_Y(d, b^{\otimes n})$  est maximale si

II.43. 
$$d_0(Y) = 1 \text{ pour tout } Y \text{ tel que } Ah_0(Y) > \int h_\nu(Y, \Delta m) d\eta(\Delta m)$$
  
 pour  $\nu = 1 \dots m$   
 $d_\nu(Y) = 1$  pour tout  $Y$  et  $\nu$  tels que  

$$\int h_\nu(Y, \Delta m) d\eta(\Delta m) > A h_0(Y)$$
  

$$\int h_\nu(Y, \Delta m) d\eta(\Delta m) > \int h_\xi(Y, \Delta m) d\eta(\Delta m)$$
  
 pour  $\xi = 1 \dots m$   
 $\xi \neq \nu$

sauf peut-être sur un ensemble de  $Y$  de mesure nulle. En revenant en  $\underline{x}$ , compte tenu de II.42. on obtient la caractérisation II.41.

Réciproquement une règle de décision satisfaisant aux hypothèses du corollaire II.16. et définie presque partout par II.43 est constante presque partout sur les orbites associées au groupe  $G_3$  donc invariante sous  $G_3$ . D'autre part elle maximise

$I_Y(d, b_{0n})$  donc  $I(d, b_{0n})$ . En conséquence  $d$  est une règle double de Bayes relativement à la probabilité a priori  $b_{0n}$ , et invariante sous  $G_3$ .

II.44. Remarque.

En définitive, on a montré que maximiser l'intégrale  $I(d, b_{0n})$  sous les conditions d'invariance équivaut à maximiser l'intégrale  $I_Y(d, b_{0n})$  sans conditions.

### CHAPITRE III

#### TESTS OPTIMAUX POUR PLUSIEURS VALEURS ABERRANTES DE MATRICE DES VARIANCES-COVARIANCES $\Lambda$ CONNUE

##### §0 INTRODUCTION ET NOTATION

Le but de ce chapitre est de généraliser à plusieurs dimensions des résultats connus en dimension 1 (voir Mac Millan, R.G. (1968)) et de présenter certains tests dans le cas de plusieurs valeurs aberrantes multidimensionnelles, en démontrant leur optimalité.

Nous supposons que  $x_1 \dots x_n$  est un  $n$  - échantillon gaussien multidimensionnel d'une loi  $\mathcal{N}_p(m, \Lambda)$  à valeurs dans  $\mathbb{R}^p$  de moyenne  $m$  inconnue et de matrice des variances-covariances  $\Lambda$  connue. Notre étude se situe dans le cadre du modèle  $\mathcal{A}$ . C'est-à-dire que nous supposons qu'une valeur aberrante provient d'un glissement en moyenne, soit d'une loi  $\mathcal{N}_p(m + \Delta m, \Lambda)$ ,  $\Delta m \in \mathbb{R}^p$ ,  $\Delta m \neq 0$ .

L'autre modèle, appelé modèle  $\mathcal{B}$ , consiste à supposer qu'une valeur aberrante provient d'une dilatation de la matrice des variances-covariances soit d'une loi  $\mathcal{N}_p(m, \lambda \Lambda)$  avec  $\lambda$  réel  $> 1$ . Il est certain que la réalité est plus complexe et qu'il faudrait par exemple combiner les deux modèles pour décrire de façon plus exacte le problème des valeurs aberrantes.

Kudo, A. (1956) a fait un pas en ce sens. Il n'en reste pas moins que, pris séparément, ces modèles nous permettent d'étudier de façon commode la puissance de nos tests.

Nous utilisons des résultats démontrés au chapitre II. Nous supposons, dans un premier temps les glissements colinéaires. Dans un deuxième temps nous les supposons

orthogonaux deux à deux relativement au produit scalaire associé à  $\Lambda^{-1}$ .

### §1 GLISSEMENTS COLINEAIRES

#### Formulation du problème et théorème principal.

Soient  $x_1 \dots x_n$   $n$  variables aléatoires gaussiennes à valeurs dans  $\mathbb{R}^p$  de moyenne  $m_1, \dots, m_n$  respectivement et de même matrice de variance-covariances  $\Lambda$  connue. Nous voulons tester l'hypothèse  $H_0 = H(m_1 = \dots = m_n = m)$  contre les hypothèses alternatives que  $k \leq n$  moyennes des observations ont glissé suivant des vecteurs de  $\mathbb{R}^p$  colinéaires à un même vecteur  $\Delta m \in \mathbb{R}^p$ ,  $k$  étant connu. En général on suppose même, pour des raisons évidentes :  $k < n/2$ . Soit  $\sigma$  une permutation de l'ensemble  $\{1, \dots, n\}$ , l'hypothèse  $H_\sigma$  est que

$$(III.1) \quad \begin{cases} E(x_1) = \alpha_{\sigma(1)} \Delta m + m \\ E(x_2) = \alpha_{\sigma(2)} \Delta m + m \\ \vdots \\ E(x_n) = \alpha_{\sigma(n)} \Delta m + m \end{cases}$$

où  $\alpha_1 \dots \alpha_n$  sont des scalaires réels connus. Nous avons donc  $n!$  hypothèses alternatives. Les  $\alpha_p$  ne sont pas nécessairement distincts et certains peuvent être nuls, l'un au moins étant différent de zéro.

Nous formulons notre problème comme un problème de décision multiple, tout en particularisant  $H_0$ .

Nous noterons  $\mathcal{G}$  le groupe des permutations de l'ensemble  $\{1, \dots, n\}$ . Soit  $D_0$  la décision d'agir comme si  $H_0$  était vraie et  $D_\sigma$  la décision d'agir comme si  $H_\sigma$  était vérifiée. Soit  $d(x_1, \dots, x_n)$  une règle de décision. (voir chapitre II Introduction 5°)

$$d(x_1, \dots, x_n) = \begin{pmatrix} d_0(x_1, \dots, x_n) \\ d_{\sigma_1}(x_1, \dots, x_n) \\ \vdots \\ d_{\sigma_{n!}}(x_1, \dots, x_n) \end{pmatrix}$$

où 
$$d_0(x_1, \dots, x_n) + \sum_{\sigma \in \mathcal{G}} d_\sigma(x_1 \dots x_n) = 1$$

$$d_0 = P(D_0 / x_1, \dots, x_n)$$

$$d_\sigma = P(D_\sigma / x_1, \dots, x_n)$$

Nous formulons les hypothèses supplémentaires ci-dessous. Les règles de décision auxquelles nous nous limitons doivent vérifier les conditions suivantes :

III.2. 1° La probabilité de choisir  $D_0$  sous l'hypothèse  $H_0$  est

$$P(D_0/d, H_0) = 1 - \alpha$$
 où  $\alpha$  est le niveau de signification du test, donc  $\alpha$  fixé, indépendant de  $d$ .

2° Les règles de décision sont invariantes lorsqu'on ajoute un vecteur constant  $y \in \mathbb{R}^p$  à chaque observation :

$$\forall y \in \mathbb{R}^p, \forall \sigma \in \mathcal{G} \quad d_\sigma(x_1 + y, \dots, x_n + y) = d_\sigma(x_1 \dots x_n)$$

et  $d_0(x_1 + y, \dots, x_n + y) = d_0(x_1 \dots x_n)$

Les règles de décisions auxquelles on se limite sont donc invariantes sous le groupe  $G_1$

3° Enfin, les règles de décision doivent vérifier : les probabilités de décision  $P(D_\sigma/d, (H_\sigma, \Delta_m))$  sont indépendantes de  $\sigma \in \mathcal{G}$ .

### III.3. Théorème.

Dans le cadre du modèle  $\mathcal{H}$ , décrit ci-dessus, la règle de décision satisfaisant aux hypothèses 1-2-3 qui maximise les probabilités de décision

$P(D_\sigma/d, (H_\sigma, \Delta_m))$  uniformément en  $\Delta_m \in \mathbb{R}^p$  et  $\sigma \in \mathcal{G}$ , nous est donnée par :

$$\left| \begin{array}{ll} \text{si } R^2_{(\sigma)} = \max_{\sigma \in \mathcal{G}} R^2_{\sigma} < \lambda_a & \text{prendre la décision } D_0 \\ \text{si } R^2_{(\sigma)} > \lambda_a & \text{prendre la décision } D_{(\sigma)} \end{array} \right.$$

où  $(\sigma)$  est l'indice pour lequel  $R^2_{\sigma}$  est maximal,  $\lambda_a$  est une constante qui dépend du niveau de signification et

$$R^2_{\sigma} = \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu} - \bar{x}) \right]' \Lambda^{-1} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu} - \bar{x}) \right]$$

III.4. Remarque.

La probabilité  $P(D_\sigma/d, (H_\sigma, \Delta m))$  étant indépendante de  $\sigma \in \mathcal{G}$ , on l'appelle probabilité de rejet de  $k$  valeurs aberrantes et l'on dit que la règle du théorème III.3. maximise uniformément en  $\Delta m$  la probabilité de rejet de  $k$  valeurs aberrantes éventuelles.

(Preuve du théorème : les probabilités de décision d'une règle de décision vérifiant l'invariance sous le groupe  $G_1$  sont indépendants de  $m$ , la moyenne inconnue. En d'autres termes, la performance des règles de décision invariantes sous  $G_1$  ne dépend pas de  $m$ . Nous supposons donc  $m = 0$ .

Nous allons rechercher la règle double de Bayes relative à une famille de probabilités a priori  $\{\eta_C/C \in \mathbb{R}_*^+\}$  définies sur le paramètre  $\Delta m \in \mathbb{R}^P$ , en prenant pour fonction de coût la fonction  $L_0$  définie au chapitre II §2, et pour  $b$  la probabilité a priori sur  $\mathcal{H}$  définie au chapitre II §3.

Mais, avant toute chose, nous avons besoin de calculer la densité de l'invariant maximal sous le groupe  $G_1$  sous les hypothèses  $H_0$  et  $H_0, \sigma \in \mathcal{G}$ .

Ce résultat est un corollaire d'un lemme dû à Kudô, 1. (1957) dont nous donnons ici les passages importants de la démonstration.

III.5. Lemme - Kudô, A. (1957).

Soient  $x_1 \dots x_n$ ,  $n$  variables aléatoires gaussiennes à valeurs dans  $\mathbb{R}^P$  de moyennes respectives  $m_1 \dots m_n$  et de même matrice des variances-covariances  $\Lambda$ . La densité conjointe des  $(n-1)$  variables  $x_1 - x_n, \dots, x_{n-1} - x_n$  que nous notons  $h(x_1, \dots, x_n)$  nous est donnée par

$$h(x_1 \dots x_n) = C_0 \exp \left\{ - \frac{1}{2n} Q \right\}$$

$$\text{où } Q = \sum_{ij} \lambda^{ij} \left\{ n \sum_{\mu=1}^n (x_{\mu i} - m_{\mu i})(x_{\mu j} - m_{\mu j}) - \left[ \sum_{\mu=1}^n (x_{\mu i} - m_{\mu i}) \right] \left[ \sum_{\mu=1}^n (x_{\mu j} - m_{\mu j}) \right] \right\}$$

et  $(\lambda^{ij})$  est la matrice inverse de la matrice  $\Lambda$ , et où  $C_0 = \frac{1}{n^{P/2} [(2\pi)^P |\Lambda|]^{n-1}}$

(Preuve : posons

$$\begin{aligned} y_1 &= x_1 - x_n \\ \vdots & \\ y_{n-1} &= x_{n-1} - x_n \end{aligned}$$

et considérons les  $(n-1)p$  variables gaussiennes  $y_{\mu i}$  à valeurs dans  $\mathbb{R}$  pour  $\mu = 1, \dots, n-1$  et  $i = 1 \dots p$  prises dans l'ordre suivant :

$$y_{11} \dots y_{n-1,1}, y_{12}, \dots y_{n-1,2} \dots y_{1p} \dots y_{n-1,p}$$

On vérifie aisément que la matrice des variances-covariances des  $(n-1)p$  variables  $y_{\mu i}$  s'écrit

$$\Lambda \otimes M(n-1)$$

où  $\otimes$  désigne le produit de Kronecker (ou tensoriel) de 2 matrices, où  $\Lambda$  est la matrice des variances-covariances des  $x_\mu$  pour  $\mu = 1 \dots n$  et  $M(n-1)$  l'élément de  $GL(n-1, \mathbb{R})$  défini par

$$M_{\nu\mu} = 1 + \delta_{\nu\mu}$$

avec  $\delta_{\nu\mu} = 1$  si  $\mu = \nu$        $0$  si  $\mu \neq \nu$

Nous savons (voir Chambadal, L. et Ovaert, J.L. (1968) et Anderson, T.S. (1958)) que

$$(\Lambda \otimes M(n-1))^{-1} = \Lambda^{-1} \otimes M^{-1}(n-1)$$

et que  $\det \Lambda \otimes M(n-1) = (\det \Lambda)^{n-1} \cdot (\det M(n-1))^p$

On démontre facilement que  $\det M(n-1) = n$ ,

soit:  $\det \Lambda \otimes M(n-1) = |\Lambda|^{n-1} n^p$ , et que  $M^{-1}(n-1)$  nous est donné par

$$M^{-1}(n-1) = \frac{1}{n} \begin{pmatrix} n-1 & -1 & \dots & 1 \\ -1 & n-1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ -1 & 1 & \dots & n-1 \end{pmatrix}$$

Alors le calcul des probabilités nous enseigne que la densité conjointe des  $(n-1)p$  variables gaussiennes

$$y_{11} \dots y_{n-1,1} \dots y_{1p} \dots y_{n-1,p}$$

nous est donnée par

$$\frac{1}{(2\pi)^{\frac{p(n-1)}{2}} (n^p |\Lambda|^{n-1})^{1/2}} \exp \left[ -\frac{1}{2} (Y_{11} \dots Y_{n-1,p})' \Lambda^{-1} \otimes M^{-1}(n-1) \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{n-1,p} \end{pmatrix} \right]$$

où  $Y_{\mu i} = y_{\mu i} - E(y_{\mu i})$  pour  $\mu = 1, \dots, n-1$   $i = 1 \dots p$   
 où  $\Lambda^{-1} \otimes M^{-1}(n-1)$  est la matrice carrée d'ordre  $(n-1)p$  donnée par :

$$\Lambda^{-1} \otimes M^{-1}(n-1) = \frac{1}{n} \begin{pmatrix} (n-1)\lambda^{11} & -\lambda^{11} & \dots & -\lambda^{11} & & (n-1)\lambda^{1p} & -\lambda^{1p} & \dots & -\lambda^{1p} \\ -\lambda^{11} & (n-1)\lambda^{11} & \dots & -\lambda^{11} & & -\lambda^{1p} & (n-1)\lambda^{1p} & \dots & -\lambda^{1p} \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ -\lambda^{11} & -\lambda^{11} & \dots & (n-1)\lambda^{11} & & -\lambda^{1p} & -\lambda^{1p} & \dots & (n-1)\lambda^{1p} \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ (n-1)\lambda^{p1} & -\lambda^{p1} & \dots & -\lambda^{p1} & & (n-1)\lambda^{pp} & -\lambda^{pp} & \dots & -\lambda^{pp} \\ -\lambda^{p1} & (n-1)\lambda^{p1} & \dots & -\lambda^{p1} & & -\lambda^{pp} & (n-1)\lambda^{pp} & \dots & -\lambda^{pp} \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ -\lambda^{p1} & -\lambda^{p1} & \dots & (n-1)\lambda^{p1} & & -\lambda^{pp} & -\lambda^{pp} & \dots & (n-1)\lambda^{pp} \end{pmatrix}$$

En effectuant le produit entre crochets et revenant en  $x_1 \dots x_n$ , après de nombreux calculs, on trouve l'expression du lemme).

III. 6. Corollaire.

Sous l'hypothèse  $H_0$  la densité de l'invariant maximal  $x_1 - x_n, \dots, x_{n-1}$  nous est donnée par

$$h_0(x_1 \dots x_n) = C_0 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (\Delta x_{\mu})' \Lambda^{-1} (\Delta x_{\mu}) \right\}$$

où  $\Delta x_{\mu} = x_{\mu} - \bar{x}$

(Preuve : par hypothèse  $m_1 = \dots = m_n = 0$ )

$$Q = \sum_{ij} \lambda^{ij} \left\{ n \sum_{\mu=1}^n x_{\mu i} x_{\mu j} - n^2 \bar{x}_i \bar{x}_j \right\}$$

où  $\bar{x}_i = \frac{1}{n} \sum_{\mu=1}^n x_{\mu i}$

$$Q = \sum_{ij} \lambda^{ij} \left\{ n \sum_{\mu=1}^n (x_{\mu i} - \bar{x}_i)(x_{\mu j} - \bar{x}_j) \right\}$$

$$\begin{aligned}
 &= n \sum_{ij=1}^p \lambda^{ij} \sum_{\mu=1}^n (\Delta x_{\mu})_i (\Delta x_{\mu})_j \\
 &= n \sum_{\mu=1}^n \sum_{i,j=1}^p \lambda^{ij} (\Delta x_{\mu})_i (\Delta x_{\mu})_j = n \sum_{\mu=1}^n (\Delta x_{\mu})' \Lambda^{-1} (\Delta x_{\mu})
 \end{aligned}$$

d'où le résultat).

### III.7. Corollaire.

Sous l'hypothèse  $H_{\sigma}$ ,  $\sigma \in \mathcal{G}$  définie en III.1. la densité de l'invariant maximal  $x_1 - x_n, \dots, x_{n-1} - x_n$  s'écrit

$$h_{\sigma}(x_1, \dots, x_n, \Delta m) = h_0 \exp \left\{ c_1 \frac{\|\Delta m\|^2}{\Lambda^{-1}} + \langle \Delta m, \sum_{\mu=1}^n \alpha_{\sigma(\mu)} \Delta x_{\mu} \rangle_{\Lambda^{-1}} \right\}$$

où  $\langle \cdot, \cdot \rangle_{\Lambda^{-1}}$  est le produit scalaire associé à  $\Lambda^{-1}$  et  $\|\cdot\|_{\Lambda^{-1}}$  la norme associée à ce produit scalaire, où

$$c_1 = \frac{+1}{2n} \left[ \left( \sum_{\mu=1}^n \alpha_{\sigma(\mu)} \right)^2 - n \left( \sum_{\mu=1}^n \alpha_{\sigma(\mu)}^2 \right) \right]$$

(Preuve : par hypothèse pour  $\mu = 1 \dots n$

$$m_{\mu} = \alpha_{\sigma(\mu)} \Delta m$$

Q s'écrit :

$$\begin{aligned}
 Q &= \sum_{ij} \lambda^{ij} \left\{ n \sum_{\mu=1}^n (x_{\mu i} - \alpha_{\sigma(\mu)} \Delta m_i) (x_{\mu j} - \alpha_{\sigma(\mu)} \Delta m_j) \right. \\
 &\quad \left. - \left[ \sum_{\mu=1}^n (x_{\mu i} - \alpha_{\sigma(\mu)} \Delta m_i) \right] \left[ \sum_{\mu=1}^n (x_{\mu j} - \alpha_{\sigma(\mu)} \Delta m_j) \right] \right\} \\
 &= \sum_{ij} \lambda^{ij} \left\{ n \sum_{\mu=1}^n x_{\mu i} x_{\mu j} - n \Delta m_j \sum_{\mu=1}^n \alpha_{\sigma(\mu)} x_{\mu i} \right. \\
 &\quad \left. - n \Delta m_i \sum_{\mu=1}^n \alpha_{\sigma(\mu)} x_{\mu j} + n \Delta m_i \Delta m_j \sum_{\mu=1}^n \alpha_{\sigma(\mu)}^2 \right\} \\
 &= n^2 \bar{x}_i \bar{x}_j + n \bar{x}_j \Delta m_i \sum_{\mu=1}^n \alpha_{\sigma(\mu)} + n \bar{x}_i \Delta m_j \sum_{\mu=1}^n \alpha_{\sigma(\mu)} \\
 &\quad - \left( \sum_{\mu=1}^n \alpha_{\sigma(\mu)} \right)^2 \Delta m_i \Delta m_j
 \end{aligned}$$

$$h_{\sigma}(x_1, \dots, x_n) = h_0 \exp \left\{ - \frac{1}{2n} \left[ \sum_{ij} \lambda^{ij} \Delta m_i \Delta m_j \left( n \sum_{\mu=1}^n \alpha_{\sigma(\mu)}^2 - \left( \sum_{\mu=1}^n \alpha_{\sigma(\mu)} \right)^2 \right) \right] \right. \\ \left. + \frac{1}{2n} \left[ 2n \sum_{ij} \lambda^{ij} \Delta m_i \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu i} - \bar{x}_i) \right] \right\}$$

en tenant compte de nos notations

$$Q = h_0 \exp \left\{ C_1 \Delta m' \Lambda^{-1} + \Delta m' \Lambda^{-1} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu i} - \bar{x}_i) \right] \right\}$$

on en conclut le corollaire).

Nous définissons sur  $\mathcal{H}$  la probabilité a-priori  $b$  telle que

$$b[H_0] = b_0 \\ b[H_{\sigma}] = \frac{1 - b_0}{n!}$$

Soit  $C$  une constante fixée telle que  $||\Delta m||_{\Lambda^{-1}}^2 = c$

Nous définissons la probabilité a priori  $\eta_c(\Delta m)$  comme étant la mesure concentrée sur l'ellipsoïde  $||\Delta m||_{\Lambda^{-1}}^2 = c$  telle qu'en le transformant par une application linéaire en la sphère unité  $S$  de  $\mathbb{R}^p$  la mesure image obtenue soit la mesure uniforme  $dS(\Delta m)$  sur la sphère unité. Nous prendrons comme fonction de coûts la fonction  $L_0$  définie chapitre II §2 et recherchons les règles doubles de Bayes relatives à la probabilité a priori  $b \otimes \eta_c$  sur  $\mathcal{H} \times \mathbb{R}^p$ . Alors en utilisant la caractérisation II.41, nous savons que la règle double de Bayes (unique à une équivalence près) nous est donnée par

$$d_{\sigma}(x) = 1 \quad \text{si } Ah_0 > \int h_{\sigma}(x_1, \dots, x_n, \Delta m) d\eta_c(\Delta m) \quad \forall \sigma \in \mathcal{C} \\ d_{\sigma'}(x) = 1 \quad \text{si } \int h_{\sigma}(x_1, \dots, x_n, \Delta m) d\eta_c(\Delta m) > Ah_0 \\ \text{et } \int h_{\sigma'}(x_1, \dots, x_n, \Delta m) d\eta_c(\Delta m) > \int h_{\sigma}(x_1, \dots, x_n, \Delta m) d\eta_c(\Delta m) \\ \forall \sigma' \in \mathcal{C} \text{ avec } \sigma' \neq \sigma$$

où  $A$  est la constante  $\in \mathbb{R}^+$  qui dépend du niveau de signification  $\alpha$ ; posons

$$\text{alors} \quad Z_{\sigma} = \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu} - \bar{x})$$

Nous nous proposons de calculer

$$I_{\sigma} = \int \frac{h_{\sigma}(x_1, \dots, x_n, \Delta m) d\eta_c(\Delta m)}{||\Delta m||_{\Lambda^{-1}}^2}$$

où  $h_\sigma$  nous est donnée par le corollaire III.7.

$$I_\sigma = h_0 \int_{\|\Delta\|^2_{\Lambda^{-1}}} \exp \{ c_1 \|\Delta m\|^2_{\Lambda^{-1}} + \langle \Delta m, Z_\sigma \rangle_{\Lambda^{-1}} \} d\eta_C(\Delta m)$$

Soit  $D$  la matrice associée à l'application linéaire qui transforme l'ellipsoïde  $\|\Delta m\|^2_{\Lambda^{-1}} = c$  en la sphère unité  $\|\Delta m\|^2 = 1$  où  $\|\cdot\|$  désigne la norme euclidienne de  $(\mathbb{R}^p)$ . Si  $J_p$  désigne la matrice identité d'ordre  $p$  nous avons

$$J_p = D' \Lambda^{-1} D \iff \Lambda^{-1} = (D')^{-1} D^{-1}$$

Dans l'intégrale opérons le changement de variable  $\Delta m \rightarrow D \Delta m$  et appliquons le résultat sur la mesure image. L'intégrale devient

$$I_\sigma = h_0 e^{c_1} \int_{\|\Delta m\|^2 = 1} \exp \{ \langle \Delta m, D^{-1} Z_\sigma \rangle \} dS(\Delta m)$$

Opérons alors un changement de base orthogonale de telle sorte que dans le nouveau système de coordonnées le vecteur  $D^{-1} Z_\sigma$  aient ses  $(p-1)$  premières composantes nulles et la  $p^{\circ}$  strictement positive. La distribution  $dS(\Delta m)$  étant uniforme nous avons

$$\begin{aligned} I_\sigma &= h_0 e^{c_1} \int_{\|\Delta m\|^2 = 1} \exp(\Delta m_p \|D^{-1} Z_\sigma\|) dS(\Delta m) \\ &= h_0 e^{c_1} \int_{\substack{\|\Delta m\|^2 = 1 \\ \Delta m_p > 0}} \exp(\Delta m_p \|D^{-1} Z_\sigma\|) dS(\Delta m) + h_0 e^{c_1} \int_{\substack{\|\Delta m\|^2 = 1 \\ \Delta m_p > 0}} \exp(-\Delta m_p \|D^{-1} Z_\sigma\|) dS(\Delta m) \\ &= 2h_0 e^{c_1} \int_{\substack{\|\Delta m\|^2 = 1 \\ \Delta m_p > 0}} \text{ch}(\Delta m_p \|D^{-1} Z_\sigma\|) dS(\Delta m) \end{aligned}$$

pour  $\Delta m_p > 0$ ,  $\text{ch}(\Delta m_p \|D^{-1} Z_\sigma\|)$  est une fonction strictement croissante de  $\|D^{-1} Z_\sigma\|$ .

Il en est de même de  $I_\sigma$  et l'on a

$$Ah_0 > I_\sigma \iff A > 2e^{c_1} \int_{\substack{\|\Delta m\|^2 = 1 \\ \Delta m_p > 0}} \text{ch}(\Delta m_p \|D^{-1} Z_\sigma\|) dS(\Delta m)$$

$$\Leftrightarrow \|D^{-1}Z_{\sigma}\| < \lambda_a \quad \text{pour tout } \sigma \in \mathfrak{G}$$

Choisir  $\lambda_a$  entre 0 et  $1^\infty$  de telle sorte que l'hypothèse 3 soit vérifiée revient à déterminer  $A \in \mathfrak{R}^*$  et par conséquence  $b_0 = b[H_0]$ .

D'autre part pour  $\sigma$  et  $\sigma' \in \mathfrak{G}$

$$I_{\sigma} > I_{\sigma'} \Leftrightarrow \int_{\substack{\|\Delta m\|^2=1 \\ \Delta m_p > 0}} \text{ch}(\Delta m_p \|D^{-1}Z_{\sigma}\|) dS(\Delta m) > \int_{\substack{\|\Delta m\|^2=1 \\ \Delta m_p > 0}} \text{ch}(\Delta m_p \|D^{-1}Z_{\sigma'}\|) dS(\Delta m)$$

$$\Leftrightarrow \|D^{-1}Z_{\sigma}\| > \|D^{-1}Z_{\sigma'}\|$$

Remarquons que  $\|D^{-1}Z_{\sigma}\| = Z'_{\sigma} (D^{-1})' D^{-1} Z_{\sigma} = Z'_{\sigma} (D')^{-1} D^{-1} Z_{\sigma}$

$$= Z'_{\sigma} \Lambda^{-1} Z_{\sigma}$$

relativement à la distribution a priori  $\eta_c(\Delta m)$  les règles doubles de Bayes symétriques invariantes sous  $G_1$  sont caractérisées par

$$d_0(x) = 1 \quad \text{si}$$

$$R_{(\sigma)}^{\lambda} = \max_{\sigma \in \mathfrak{G}} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu} - \bar{x}) \right]' \Lambda^{-1} \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu} - \bar{x}) < \lambda_a$$

$$d_{\sigma}(x) = 1 \quad \text{si } R_{(\sigma)} > \lambda_a$$

Cette règle de décision est manifestement symétrique en  $x_1 \dots x_n$ . Les densités des lois conjointes des variables  $x_1 \dots x_n$  sous les diverses hypothèses étant également symétriques en  $x_1 \dots x_n$ , on en conclut que la condition 3° est vérifiée.

Pour terminer la démonstration du théorème nous avons besoin du résultat suivant :

### III.8. Lemme.

Soit  $d$  la règle de décision ci-dessus, alors  $\forall \sigma \in \mathfrak{G}$ , les probabilités de décision  $P(D_{\sigma}/d, (H_{\sigma}, \Delta m))$  ne dépendent que de  $\|\Delta m\|_{\Lambda^{-1}}$

(Preuve : Soit

$$d = (d_0, d_{\sigma_1}, \dots, d_{\sigma_n!})$$

la règle de décision définie ci-dessus.

Alors

$$P(D_\sigma/d, (H_\sigma, \Delta m)) = \int_{\mathbb{R}^{n.p}} d_\sigma(x_1 \dots x_n) f_\sigma(x_1 \dots x_n, \Delta m) dx$$

où  $f_\sigma(x_1 \dots x_n, \Delta m)$  est la densité conjointe des  $n$  variables aléatoires  $X_1, \dots, X_n$  sous l'hypothèse  $H_\sigma, \Delta m$

On trouve facilement, les variables  $X_\mu$ , pour  $\mu = 1 \dots n$  étant indépendantes, que

$$f_\sigma(x_1 \dots x_n, \Delta m) = \frac{1}{(2\pi)^{np/2} |\Lambda|^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (x_\mu - \alpha_{\sigma(\mu)} \Delta m)' \Lambda^{-1} (x_\mu - \alpha_{\sigma(\mu)} \Delta m) \right\}$$

posons

$$c_2 = \frac{1}{(2\pi)^{np/2} |\Lambda|^{n/2}}$$

il vient

$$c_2 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n x_\mu' \Lambda^{-1} x_\mu + \left( \sum_{\mu=1}^n \alpha_{\sigma(\mu)} x_\mu \right)' \Lambda^{-1} \Delta m - \frac{1}{2} \sum_{\mu=1}^n \alpha_{\sigma(\mu)}^2 \|\Delta m\|_{\Lambda^{-1}}^2 \right\}$$

$$P(D_\sigma/d, (H_\sigma, \Delta m)) = c_2 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n \alpha_{\sigma(\mu)}^2 \|\Delta m\|_{\Lambda^{-1}}^2 \right\}$$

$$\int_{\mathbb{R}^{np}} d_\sigma(x_1 \dots x_n) \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n x_\mu' \Lambda^{-1} x_\mu + Z_\sigma' \Lambda^{-1} \Delta m \right\} dx$$

$$= c_2 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n \alpha_{\sigma(\mu)}^2 \|\Delta m\|_{\Lambda^{-1}}^2 \right\}$$

$$\int_{\mathbb{R}^{n.p}} d_\sigma(x_1 \dots x_n) \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n x_\mu' \Lambda^{-1} x_\mu + \langle D^{-1} Z_\sigma, D^{-1} m \rangle \right\} dx$$

où  $\langle \cdot, \cdot \rangle$  désigne le produit scalaire ordinaire de  $\mathbb{R}^p$  et  $D$  la matrice telle que

$$D' \Lambda^{-1} D = \mathbb{1}_p$$

en opérant un changement de base orthogonale de telle sorte que  $D^{-1} \Delta m$  ait ses (p-1) premières composantes nulles il vient

$$P(D_\sigma/d, (H_\sigma, \Delta m)) = c_2 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n \alpha^2_{\sigma(\mu)} \|\Delta m\|_{\Lambda^{-1}}^2 \right\}$$

$$\int_{\mathbb{R}^{np}} d_\sigma(x_1 \dots x_n) \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n x'_\mu \Lambda^{-1} x_\mu + \|D^{-1} \Delta m\| \right.$$

$$\left. P(D^{-1} Z_\sigma)_p \, d\underline{x} \right.$$

où  $\|D^{-1} \Delta m\| = \|\Delta m\|_{\Lambda^{-1}}$ ,  $P$  est la matrice de passage

et  $(.)_p$  désigne la  $p^{\circ}$  composante d'un vecteur de  $\mathbb{R}^p$ ; d'où le résultat.)

On peut alors appliquer le théorème II.20. La démonstration du théorème III.3. est alors complète).

## §2 EXEMPLES D'APPLICATIONS

Nous donnons ici trois corollaires qui nous permettent de retrouver le résultat connu de Kudo, A. (1957) et de généraliser certains résultats en dimension 1, présentés par Mac Millan, R.G. (1968).

### III.9. Corollaire.

Sous l'hypothèse qu'il n'y a au plus qu'une valeur aberrante, dans le cadre du modèle  $\mathcal{H}$  décrit au paragraphe 1, la règle de décision satisfaisant aux hypothèses 1.2.3. qui maximise les probabilités de décision  $P_{H_\nu, \Delta m}(D_\nu)$  pour  $\nu = 1 \dots n$  uniformément en  $\Delta m \in \mathbb{R}^p$  nous est donnée par :

$$\left| \begin{array}{l} \text{prendre la décision } D_0 \text{ si } \chi^2_{(n)} = \max_{\mu=1 \dots n} (x_\mu - \bar{x})' \Lambda^{-1} (x_\mu - \bar{x}) < \lambda_a \\ \text{prendre la décision } D_{(n)} \text{ si } \chi^2_{(n)} > \lambda_a \end{array} \right.$$

(Preuve : dans ce cas, un seul des  $\alpha_\mu$  est non nul, égal à 1 et il y a  $n$  hypothèses alternatives).

Ce résultat était déjà connu et dû à Kudo, A.

III.10. Corollaire.

Supposons que  $1 < k < n$  des  $\alpha_\mu$  soient non nuls, tous égaux à 1.

Alors la règle de décision satisfaisant aux hypothèses du théorème III.3. et possédant les mêmes propriétés nous est donnée par :

$$\left| \begin{array}{l} \text{prendre la décision } D_0 \text{ si } R^2_{(\sigma)} < \lambda_a \\ \text{prendre la décision } D_{(\sigma)} \text{ si } R^2_{(\sigma)} > \lambda_a \end{array} \right.$$

où  $R^2_{(\sigma)} = \text{Max}_{\sigma \in \mathcal{C}} \left[ \sum_{\mu=1}^k (x_{\sigma(\mu)} - \bar{x}) \right]' \Lambda^{-1} \left[ \sum_{\mu=1}^k (x_{\sigma(\mu)} - \bar{x}) \right]$  et  $\lambda_a$  est une constante  
dépendant de a le niveau de signification.

(Preuve : sans perdre de généralité, nous pourrions supposer

$$\alpha_1 = \dots = \alpha_k = 1$$

$$\alpha_{k+1} = \dots = \alpha_n = 0$$

alors  $\sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_\mu - \bar{x}) = \sum_{\mu=1}^n \alpha_\mu (x_{\sigma^{-1}(\mu)} - \bar{x})$

Alors  $\text{Max}_{\sigma \in \mathcal{C}} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_\mu - \bar{x}) \right]' \Lambda^{-1} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_\mu - \bar{x}) \right]$

$$= \max_{\sigma^{-1} \in \mathcal{C}} \left[ \sum_{\mu=1}^n \alpha_\mu (x_{\sigma^{-1}(\mu)} - \bar{x}) \right]' \Lambda^{-1} \left[ \sum_{\mu=1}^n \alpha_\mu (x_{\sigma^{-1}(\mu)} - \bar{x}) \right]$$

$$= \max_{\sigma \in \mathcal{C}} \left[ \sum_{\mu=1}^n \alpha_\mu (x_{\sigma(\mu)} - \bar{x}) \right]' \Lambda^{-1} \left[ \sum_{\mu=1}^n \alpha_\mu (x_{\sigma(\mu)} - \bar{x}) \right]$$

en remplaçant les  $\alpha_\mu$  par leur valeur on obtient le résultat.)

Supposons que les  $n$  observations soient partagées en  $r$  populations de  $k$  éléments :

$$n = rk$$

Nous nous proposons de savoir si l'une des populations a glissé en moyenne.

III.11. Corollaire.

La règle de décision correspondante, vérifiant les hypothèses du théorème III.3. et possédant les propriétés qui y sont énoncées, nous est donnée par

$$\text{prendre la décision } D_0 \quad \text{si } R^2_{(\ell)} < \lambda'_a$$

$$D_{(\ell)} \quad \text{si } R^2_{(\ell)} > \lambda'_a$$

où  $(\ell)$  désigne l'indice pour lequel

$$R^2_{\ell} = (\bar{x}(\ell) - \bar{x})' \Lambda^{-1} (\bar{x}(\ell) - \bar{x}) \quad \text{est maximal ; } 1 \leq \ell \leq r$$

$$\lambda'_a = \lambda_a / k^2$$

(Preuve : par hypothèse  $k$  des coefficients sont non nuls, égaux à 1, les autres étant nuls.

Les seules permutations qui présentent un intérêt sont celles pour lesquelles les coefficients  $\alpha_{\mu} \neq 0$  opèrent sur les éléments d'un même groupe. Nous avons donc  $r$  hypothèses alternatives

$$R^2_{\sigma(\ell)} = \left[ \sum_{j=1}^k (x_{\mu(j)} - \bar{x}) \right]' \Lambda^{-1} \left[ \sum_{j=1}^k (x_{\mu(j)} - \bar{x}) \right]$$

où  $\mu(j)$  désigne un des  $k$  indices  $\mu$  pour lesquels  $\alpha_{\mu} = 1$ , et  $\sigma(\ell)$  désigne la  $\ell^{\circ}$  des  $r$  permutations présentant un intérêt.

$$R^2_{\sigma(\ell)} = \left[ \sum_{j=1}^k x_{\mu(j)} - k\bar{x} \right] \Lambda^{-1} \left[ \sum_{j=1}^k x_{\mu(j)} - k\bar{x} \right]$$

$$= k^2 (\bar{x}(\ell) - \bar{x})' \Lambda^{-1} (\bar{x}(\ell) - \bar{x})$$

où

$$\bar{x}(\ell) = \frac{1}{k} \sum_{j=1}^k x_{\mu(j)}$$

d'où le résultat.)

On trouve une généralisation, lorsque  $\Lambda$  est connue, du résultat de Paulson, E.(1952) en dimension 1 qui supposait, lui, inconnu

Bien que l'on ne connaisse pas  $\Delta\pi$ , on a supposé les  $\alpha_{\mu}$  connus. Nous pouvons donc déjà dire qu'en l'absence de toute information sur la nature des glissements il n'y a pas de règle optimale de rejet de  $k > 1$  valeurs aberrantes.

§3 GLISSEMENTS ORTHOGONAUX

Lorsque les glissements sont quelconques, non colinéaires, il ne parait pas possible de trouver une règle de décision optimale qui ne soit pas liée à la connaissance préalable de ces glissements. Cependant lorsque les glissements sont deux à deux orthogonaux au sens du produit scalaire associé à l'inverse de la matrice des variances-covariances nous proposons le résultat ci-dessous.

Nous envisageons le cas de deux valeurs aberrantes éventuelles et signalons que les résultats obtenus se généralisent aisément à  $k > 2$  valeurs aberrantes. Soient  $X_1 \dots X_n$ ,  $n$  variables aléatoires gaussiennes à valeurs dans  $\mathbb{R}^P$  de moyennes respectives  $m_1 \dots m_n$  inconnues et de même matrice des variances-covariances  $\Lambda$  supposée connue. Nous voulons tester l'hypothèse  $H_0 = H(m_1 = \dots = m_n = m)$  contre les hypothèses alternatives que nous noterons  $H_{\nu\xi}$  que les valeurs  $x_\nu$  et  $x_\xi$  ont glissé en moyenne d'un vecteur  $\Delta m_1$  et  $\Delta m_2$  respectivement.

Nous supposons inconnus les vecteurs  $\Delta m_1$  et  $\Delta m_2$ , mais orthogonaux au sens  $\langle \Delta m_1, \Delta m_2 \rangle_{\Lambda^{-1}} = 0$

Le problème est donc symétrique en  $\Delta m_1$  et  $\Delta m_2$

En conséquence nous supposerons  $1 \leq \nu < \xi \leq n$ . Nous avons donc  $\frac{n(n-1)}{2}$  hypothèses alternatives  $H_{\nu\xi}$

III.12. 
$$H_{\nu\xi} = H(m_\nu = m + \Delta m_1, m_\xi = m + \Delta m_2, \nu < \xi)$$
  
$$m_\mu = m, \text{ pour } \mu = 1 \dots n, \nu \neq \mu \neq \xi)$$

Nous formulons les hypothèses 1-2-3 du §1, en respectant nos nouvelles notations. Sous ces hypothèses nous appellerons  $h_{\nu\xi}$  la densité de l'invariant maximal sous le groupe  $G_1$ , sous l'hypothèse  $H_{\nu\xi}$ . La performance des règles de décision invariantes sous  $G_1$  ne dépendant pas de  $m$  nous supposerons  $m = 0$ . La densité  $h_0(x_1, \dots, x_n)$  nous est donnée par le corollaire III.7., la densité  $h_{\nu\xi}(x_1 \dots x_n)$  nous est fournie par le résultat suivant :

III.13. Lemme.

Sous l'hypothèse  $H_{\nu\xi}$  la densité de l'invariant maximal sous le groupe  $G_1$  vaut

$$h_{\nu\xi}(x_1 \dots x_n) = h_0 \exp \left\{ -\frac{n-1}{2n} \left( \|\Delta m_1\|_{\Lambda^{-1}}^2 + \|\Delta m_2\|_{\Lambda^{-1}}^2 \right) + \langle \Delta m_1, \Delta x_{\nu} \rangle_{\Lambda^{-1}} + \langle \Delta m_2, \Delta x_{\xi} \rangle_{\Lambda^{-1}} \right\}$$

où  $\Delta x_{\nu} = x_{\nu} - \bar{x}$

(Preuve : il nous faut calculer l'expression  $-\frac{1}{2n} Q$  du lemme III.5.

avec  $m_{\nu} = \Delta m_1$ ,  $m_{\xi} = \Delta m_2$ ,  $m_{\mu} = 0$  pour  $\mu = 1 \dots n$ ,  $\nu \neq \mu \neq \xi$

En tenant compte de l'égalité :

$$\langle \Delta m_1, \Delta m_2 \rangle_{\Lambda^{-1}} = 0$$

un calcul facile nous conduit au résultat.)

Nous allons considérer cette fois que l'espace des paramètres est  $\mathbb{R}^{2p}$  soit  $k = 2p$  selon les notations du chapitre II et nous écrivons

$$\mathbb{R}^{2p} = \mathbb{R}^p \oplus \mathbb{R}^p$$

Posons  $\Delta m_S = \Delta m_1 \oplus \Delta m_2$

Le produit scalaire associé à la matrice  $\Lambda^{-1}$  et à cette somme directe que nous noterons  $\langle \cdot, \cdot \rangle_{\Lambda^{-1}}$  est défini par la matrice bloc

$$\begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix}$$

De même, on définit la norme que nous noterons  $\|\cdot\|_{\Lambda^{-1}}$  associé à ce produit scalaire.

Sur  $\Delta m_S$  nous définissons la probabilité a priori suivante :  $\eta_c$  est la probabilité concentrée sur l'ellipsoïde  $\|\Delta m_S\|_{\Lambda^{-1}} = c^{1/2}$  telle qu'en le transformant par une application linéaire en la sphère unité de  $\mathbb{R}^{2p}$  :  $\{\Delta m_S / \|\Delta m_S\| = 1\}$  la mesure image obtenue soit la mesure uniforme notée  $dS(\Delta m_S)$

Sur  $\mathcal{X}$  nous définissons la probabilité a priori suivante

$$h[H_0] = b_0$$

$$h[H_{\nu\xi}] = \frac{2(1-b_0)}{n(n-1)}$$



et  $h_{\nu\xi}(x, \Delta m_s)$  est la densité de l'invariant maximal sous le groupe  $G_1$  sous l'hypothèse  $H_{\nu\xi}$ . Pour  $\nu < \xi$  nous nous proposons de calculer l'intégrale

$$I_{\nu\xi} = \int_{\mathbb{R}^{2p}} h_{\nu\xi}(x, \Delta m_s) d\eta_c(\Delta m_s)$$

$$I_{\nu\xi} = h_0 \int_{\mathbb{R}^{2p}} \exp \left\{ -\frac{n-1}{2n} \left( \|\Delta m_s\|_{\Lambda^{-1}}^2 + \langle \Delta m_s, \Delta x_\nu \oplus \Delta x_\xi \rangle_{\Lambda^{-1}} \right) \right\} d\eta_c(\Delta m_s)$$

Soit  $D_b = \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix}$  la matrice bloc  $2p \times 2p$  associée à l'application

linéaire qui transforme l'ellipsoïde  $\|\Delta m_s\|_{\Lambda^{-1}}^2 = c$  en la sphère unité de  $\mathbb{R}^{2p}$  définie par  $\|\Delta m\|_{\Lambda^{-1}}^2 = 1$ . On a la relation

$$\begin{pmatrix} D' & 0 \\ 0 & D' \end{pmatrix} \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix} = \begin{pmatrix} \mathbb{1}_p & 0 \\ 0 & \mathbb{1}_p \end{pmatrix}$$

Dans l'intégrale opérons le changement de variable  $\Delta m_s \rightarrow D_b \Delta m_s$  et appliquons le résultat sur la mesure image. Il vient :

$$I_{\nu\xi} = h_0 e^{-\frac{n-1}{2n}} \int_{\|\Delta m_s\|=1} \exp \left\{ \langle \Delta m_s, (D')^{-1} (\Delta x_\nu \oplus \Delta x_\xi) \rangle \right\} dS(\Delta m_s)$$

Il existe un changement de base orthogonale unique de telle sorte que  $(D')^{-1} (\Delta x_\nu \oplus \Delta x_\xi)$  soit parallèle au 1<sup>er</sup> vecteur de base  $e_1$  de  $\mathbb{R}^p \oplus \mathbb{R}^p$ . Alors en opérant ce changement de base, la distribution  $dS(\Delta m_s)$  étant uniforme, il vient

$$I_{\nu\xi} = h_0 e^{-\frac{n-1}{2n}} \int_{\|\Delta m_s\|=1} \left\{ \exp (\Delta m_s)_1 \left\| D_b^{-1} (\Delta x_\nu \oplus \Delta x_\xi) \right\| \right\} dS(\Delta m_s)$$

$$= 2h_0 e^{-\frac{n-1}{2n}} \int_{\substack{\|\Delta m_s\|=1 \\ (\Delta m_s)_1 > 0}} \text{ch} \left\{ (\Delta m_s)_1 \left\| D_b^{-1} (\Delta x_\nu \oplus \Delta x_\xi) \right\| \right\} dS(\Delta m_s)$$

On en conclut que  $\int_{\nu\xi}$  est fonction croissante de

$$\left\| D_b^{-1} (\Delta x_\nu \oplus \Delta x_\xi) \right\| = (\Delta x_\nu \oplus \Delta x_\xi)' \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix} (\Delta x_\nu \oplus \Delta x_\xi) = \chi^2_\nu + \chi^2_\xi$$

Alors de la même façon qu'au paragraphe 1 on obtient

$$\int_{\nu\xi} > \int_{\mu\xi} \iff \circ ||D_b^{-1}(\Delta x_\nu \circ \Delta x_\xi)|| > \circ ||D_b^{-1}(x_\nu \circ \Delta x_\xi)||$$

et trouver  $A \in ]0, +\infty[$  tel que

$$Ah_0 > \int_{\nu\xi} \text{ pour tout } \nu \text{ et } \xi \in \{1, \dots, n\} \nu < \xi$$

est équivalent à trouver une constante  $\lambda_a \in \mathbb{R}_*^+$  telle que

$$\max_{\substack{\nu, \xi \\ \nu < \xi}} \circ ||D_b^{-1}(\Delta x_\nu \circ \Delta x_\xi)||^2 < \lambda_a$$

ce choix dépendant du niveau de signification  $\alpha$  de telle sorte que la condition 1° du §1 soit vérifiée. En définitive la règle double de Bayes relative à la probabilité a priori  $b \circ \eta_c(\Delta m_s)$  sur  $\mathcal{X} \times \mathbb{R}^{2p}$  invariante sous  $G_1$  nous est donnée par

$$d_0(\underline{x}) = 1 \text{ si } \chi_{(n)}^2 + \chi_{(n-1)}^2 < \lambda_a$$

$$d_{(n)(n-1)}(\underline{x}) = 1 \text{ si } \chi_{(n)}^2 + \chi_{(n-1)}^2 > \lambda_a$$

(n) et (n-1) désignant les indices pour lesquels on obtient les deux plus grandes valeurs de  $\chi_\nu^2$ . Cette règle est manifestement symétrique en  $x_1 \dots x_n$ . D'après la symétrie des densités des lois conjointes des  $n$  variables  $X_1 \dots X_n$  on en déduit que la condition 3° est vérifiée. Enfin comme au paragraphe 1 on démontre le résultat suivant :

### III.16. Lemme.

Pour la règle de décision  $d$  définie ci-dessus les probabilités de décision  $P(D_{\nu\xi}/d, (H_{\nu\xi}, \Delta m_1 \circ \Delta m_2))$  ne dépendent que de  $\circ ||\Delta m_1 \circ \Delta m_2||_{\Lambda^{-1}}$

En appliquant alors le théorème II.20. puis en faisant varier  $c$ , on achève la démonstration du théorème III).

#### §4 CONCLUSION

Les résultats que nous avons prouvés, nous montrent que, d'une façon générale, dans le cas de plusieurs valeurs aberrantes dans un échantillon, il n'est pas possible de trouver une règle de rejet optimale indépendante des glissements. En effet, même lorsque les glissements sont colinéaires, la règle de rejet optimale suppose connus les coefficients de proportionnalité .

Cependant si l'on suppose les glissements identiques, la règle de rejet optimale est simple. Tout se passe comme si à partir du nuage de points de  $\mathbb{R}^p$  :  $x_1 \dots x_n$ , on formait un nouveau nuage de points  $z = \frac{1}{k}(x_{\sigma(1)} + \dots + x_{\sigma(k)})$  ;  $k$  étant le nombre supposé déterminé de valeurs aberrantes. Puis on étudie l'éloignement de ces points de  $\bar{x} = \frac{1}{n} \sum_{\mu=1}^n x_{\mu}$  au sens de la distance généralisée.

De plus lorsque les glissements sont orthogonaux, le résultat présente un intérêt quant aux applications au rejet de valeurs aberrantes lors d'une analyse en composantes principale, si les glissements se font le long des axes principaux (ou presque).

## CHAPITRE IV

### TESTS POUR UN NOMBRE DETERMINE DE VALEURS ABERRANTES DE MATRICE DES VARIANCES-COVARIANCES $\Lambda$ INCONNUE

---

#### §0 INTRODUCTION ET NOTATIONS

Nous envisageons maintenant le cas où la matrice des variances-covariances  $\Lambda$  est inconnue. Nous nous situons toujours dans le cadre du modèle  $\mathcal{H}$ , où l'on suppose qu'une valeur aberrante provient d'un glissement en moyenne : soit d'une loi  $\mathcal{N}_p(m + \Delta m, \Lambda)$  où  $\Delta m$  est un vecteur non nul de  $\mathbb{R}^p$ .

Si l'on appelle  $x_1, \dots, x_n$  les  $n$  vecteurs de  $\mathbb{R}^p$  sur lesquels nous travaillons, et  $m_1, \dots, m_n$  les  $n$  moyennes des lois gaussiennes associées, nous voulons tester l'hypothèse  $m_1 = \dots = m_n$  contre un certain nombre d'hypothèses alternatives que nous allons préciser plus loin. De façon évidente notre test doit être symétrique. L'expression mathématique de la symétrie est l'invariance sous un certain groupe de transformations.

Nous reprenons les hypothèses du chapitre III §1 en y ajoutant celle, nouvelle, que nous impose l'absence d'information sur  $\Lambda$ .

#### §1 FORMULATION DU PROBLEME ET THEOREME PRINCIPAL

Soient  $X_1, \dots, X_n$   $n$  vecteurs aléatoires gaussiens de  $\mathbb{R}^p$  indépendants de moyennes respectives  $m_1 \dots m_n$  et de matrice commune des variances-covariances  $\Lambda$  ; tous les paramètres étant inconnus. Nous voulons tester l'hypothèse

$$H_0 = H(m_1 = \dots = m_n)$$

contre les hypothèses alternatives que  $k \leq n$  moyennes ont glissé suivant des vecteurs de  $\mathbb{R}^p$  colinéaires à un même vecteur  $\Delta m \in \mathbb{R}^p$ ,  $\Delta m \neq 0$ . Soit  $\sigma$  une permutation de l'ensemble  $\{1, \dots, n\}$ , l'hypothèse  $H_\sigma$  est la suivante :

$$\text{IV.1.} \quad \begin{cases} E(x_1) = \alpha_{\sigma(1)} \Delta m + m \\ E(x_2) = \alpha_{\sigma(2)} \Delta m + m \\ \vdots \\ E(x_n) = \alpha_{\sigma(n)} \Delta m + m \end{cases}$$

où  $\alpha_1, \dots, \alpha_n$  sont des scalaires réels, éventuellement nuls, supposés connus. Nous avons donc  $n!$  hypothèses alternatives. Les  $\alpha_\mu$  ne sont pas nécessairement distincts. L'un au moins, est non nul.

Notre problème est un problème de décision multiple sur une statistique maximale invariante. Nous cherchons en effet une procédure de décision symétrique qui nous permette de choisir entre les  $n! + 1$  hypothèses  $\{H_0, H_\sigma / \sigma \in \mathfrak{S}\}$  où  $\mathfrak{S}$  désigne le groupe des permutations de  $\{1, \dots, n\}$ .

Nous appellerons  $D_0$  la décision d'accepter  $H_0$

$D_\sigma$  la décision d'accepter  $H_\sigma$

Soit  $d(x_1, \dots, x_n)$  une règle de décision.

$$d(x_1, \dots, x_n) = \begin{pmatrix} d_0(x_1, \dots, x_n) \\ d_{\sigma_1}(x_1, \dots, x_n) \\ \vdots \\ d_{\sigma_{n!}}(x_1, \dots, x_n) \end{pmatrix}$$

$$\text{où} \quad \begin{cases} d_0(\underline{x}) + \sum_{\sigma \in \mathfrak{S}} d_\sigma(\underline{x}) = 1 \\ d_0(\underline{x}) \text{ est la probabilité de prendre la décision } D_0 \text{ après avoir} \\ \text{observé } (x_1, \dots, x_n) = \underline{x} \\ d_\sigma(\underline{x}) \text{ est la probabilité de prendre la décision } D_\sigma \text{ après avoir} \\ \text{observé } \underline{x} \end{cases}$$

IV.2. Nous formulons les hypothèses supplémentaires suivantes sur les règles de décisions.

1° La probabilité de choisir  $D_0$  lorsque  $H_0$  est réalisé est  $P_{H_0}(D_0) = 1 - \alpha$  où  $\alpha$  est le niveau de signification du test  $0 < \alpha < 1$  fixé et indépendant de  $d \in \mathcal{D}$

2° Les règles de décision sont invariantes lorsqu'on translate les vecteurs  $x_\mu$ ,  $\mu = 1, \dots, n$  d'un vecteur constant  $y \in \mathbb{R}^p$   $\forall y \in \mathbb{R}^p$ ,  $\forall \sigma \in \mathcal{G}$

$$d_\sigma(x_1 + y, \dots, x_n + y) = d_\sigma(x_1, \dots, x_n)$$

et  $d_0(x_1 + y, \dots, x_n + y) = d_0(x_1, \dots, x_n)$

En d'autres termes nous nous limitons aux règles invariantes sous  $G_1$

3° Les procédures de décision sont invariantes lorsqu'on multiplie les vecteurs  $x_\mu$ ,  $\mu = 1 \dots n$ , par une matrice  $(p \times p): B$ , non singulière.

$$\forall B \in GL(p, \mathbb{R}), d_0(Bx_1, \dots, Bx_n) = d_0(x_1, \dots, x_n)$$

$$\forall \sigma \in \mathcal{G} \quad d_\sigma(Bx_1, \dots, Bx_n) = d_\sigma(x_1, \dots, x_n)$$

En d'autres termes nous nous limitons aux règles invariantes sous  $G_2$

4° Les procédures de décision satisfont à :

$$P_{H_0, \Delta m}(D_\sigma) \text{ est indépendant de } \sigma \in \mathcal{G}$$

En définitive les conditions 2 et 3 font que nous nous limitons aux règles de décision invariantes sous  $G_3$  groupe engendré par  $G_1$  et  $G_2$  (voir chapitre II §6).

Nous prendrons pour fonction de coûts la fonction  $L_0$  du chapitre II §2.

Nous pouvons alors énoncer le résultat suivant.

IV.3. Théorème.

Dans le cadre de modèle  $\mathcal{A}$ , la règle de décision satisfaisant aux hypothèses 1-2-3-4 qui maximise les probabilités de décision  $P_{H_0, \Delta m}(D_\sigma)$  uniformément en  $\Delta m \in \mathbb{R}^p$  et  $\sigma \in \mathcal{G}$  nous est donnée par :

prendre la décision  $D_0$  si  $T^2_{(\sigma)} = \max_{\sigma \in \mathcal{G}} T^2_\sigma < \lambda_\alpha$

prendre la décision  $D_{(\sigma)}$  si  $T^2_{(\sigma)} > \lambda_\alpha$

où  $(\sigma)$  est l'indice pour lequel  $T^2_\sigma$  est maximal où  $\lambda_\alpha$  est une constante réelle positive dépendant de  $\alpha$  le niveau de signification du test et où

$$(IV.4.) \quad T_{\sigma}^2 = \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu} - \bar{x}) \right]' S^{-1} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_{\mu} - \bar{x}) \right]$$

$$S = \frac{1}{n-1} \sum_{\mu=1}^n (x_{\mu} - \bar{x})(x_{\mu} - \bar{x})'$$

IV. Remarque.

On dit alors que la règle ci-dessus maximise la probabilité de rejet de  $k$  valeurs aberrantes éventuelles (voir remarque III.4.)

(Preuve du théorème : les probabilités de décision d'une règle de décision invariante sous le groupe  $G_1$  ne dépendant pas de  $m$  la moyenne inconnue. Les probabilités de décision d'une règle invariante sous  $G_2$  ne dépendant pas de  $\Lambda$  la matrice des variances-covariances inconnues. En d'autres termes nous dirons que la performance d'une règle de décision invariante sous  $G_3$  (groupe engendré par  $G_1$  et  $G_2$ ) ne dépend ni de  $m$  ni de  $\Lambda$ . Pour simplifier nous supposons donc  $m = 0$ ,  $\Lambda = \int_p$  où  $\int_p$  désigne la matrice identité de  $\mathbb{R}^p$ ).

Nous allons rechercher les règles doubles de Bayes relatives à une famille de probabilité  $b_{\sigma\eta}$  sur  $\mathcal{G} \times \mathbb{R}^p$  où  $b$  est la probabilité a-priori sur  $\mathcal{G}$  définie chapitre II §3 et  $\eta$  une probabilité a-priori symétrique sur  $\Delta m \in \mathbb{R}^p$ . Nous avons besoin de calculer la densité de l'invariant maximal sous le groupe  $G_3$  (voir chapitre II §6) sous les hypothèses  $H_0$  et  $H_1$ . Pour cela nous utilisons le lemme suivant.

IV.5. Lemme.

Soient  $X_1, \dots, X_n$   $n$  variables aléatoires gaussiennes de lois respectives  $\mathcal{N}_p(m_{\mu}, \Lambda)$  pour  $\mu = 1 \dots n$ . Alors la densité de la loi conjointe des  $n-p-1$  variables  $Y_2, \dots, Y_{n-p}$  telles que

$$\begin{aligned} Y_2 &= B^{-1} Z_2 \\ &\vdots \\ Y_{n-p} &= B^{-1} Z_{n-p} \end{aligned}$$

où  $Z_{\mu} = X_{\mu} - X_1$  pour  $\mu = 1 \dots n$

et  $B$  est la matrice  $p \times p$  formée des  $p$  vecteurs colonnes  $(Z_{n-p+1}, \dots, Z_n)$ , nous est donnée par

$$h(y) = C_0 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n m'_\mu m_\mu + \frac{1}{2n} \sum_{i=1}^p \left( \sum_{\mu=1}^n m_i \right)^2 \right\}$$

$$\left[ \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (y_\mu - \bar{y})' B' B (y_\mu - \bar{y}) + \sum_{\mu=1}^n (y_\mu - \bar{y})' B m_\mu \right\} |\det B|^{n-p-1} dB \right]$$

où

$$C_1 = \frac{1}{(2\pi)^{np/2}} \left( \frac{2\pi}{n} \right)^{p/2}$$

où nous noterons

$$y = y_2, \dots, y_{n-p}$$

et

$$\bar{y} = \frac{1}{n} \sum_{\mu=1}^n y \quad \text{avec} \quad y_\mu = B^{-1} z_\mu \quad \text{pour} \quad \mu = 1 \dots n$$

(Preuve : la densité de la variable aléatoire  $X_\mu$  nous est donnée par

$$f(x_\mu) = \frac{1}{(2\pi)^{p/2}} \exp \left\{ -\frac{1}{2} (x_\mu - m_\mu)' (x_\mu - m_\mu) \right\} \quad \text{pour} \quad \mu = 1 \dots n$$

Les variables  $X_\mu$  étant indépendantes la densité de la loi conjointe  $X_1, \dots, X_n$  nous est donnée par

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{np/2}} \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (x_\mu - m_\mu)' (x_\mu - m_\mu) \right\}$$

posons

$$C_0 = 1/(2\pi)^{\frac{n \cdot p}{2}}$$

$$f(x_1, \dots, x_n) = C_0 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n x'_\mu x_\mu + \sum_{\mu=1}^n x'_\mu m_\mu - \frac{1}{2} \sum_{\mu=1}^n m'_\mu m_\mu \right\}$$

posons

$$z_\mu = x_\mu - x_1 \quad \text{pour} \quad \mu = 1 \dots n$$

$$z_1 = 0$$

$$\bar{z} = \frac{1}{n} \sum_{\mu=1}^n z_\mu$$

remarquons tout de suite que

$$\begin{aligned} \bar{z} &= \frac{1}{n} \sum_{\mu=2}^n (x_\mu - x_1) = \frac{1}{n} \sum_{\mu=2}^n x_\mu - \frac{n-1}{n} x_1 \\ &= \frac{x_d}{n} + \frac{1}{n} \sum_{\mu=2}^n x_\mu - \frac{x_1}{n} - \frac{n-1}{n} x_1 \\ &= \bar{x} - x_1 \end{aligned}$$

Nous nous proposons de calculer la densité de la loi conjointe des (n-1) variables  $Z_2, \dots, Z_n$ . Pour cela opérons la transformation

$$x_1 = w$$

$$x_\mu = w + z_\mu \quad \text{pour } \mu = 2, \dots, n.$$

Le jacobien de cette transformation vaut un. D'après un résultat connu sur la mesure image, pour obtenir la densité de la loi conjointe  $Z_2, \dots, Z_n$  nous intégrons la densité de  $w, Z_2, \dots, Z_n$  par rapport à  $w$  sur  $\mathbb{R}^p$ .

$$f(z_2, \dots, z_n) = C_0 \int_{\mathbb{R}^p} \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (w + z_\mu)' (w + z_\mu) + \sum_{\mu=1}^n (w + z_\mu)' m_\mu \right. \\ \left. - \frac{1}{2} \sum_{\mu=1}^n m_\mu' m_\mu \right\} dw$$

$$= C_0 \int_{\mathbb{R}^p} \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (w + \bar{z}_\mu - \bar{z})' (w + \bar{z}_\mu - \bar{z}) + \sum_{\mu=1}^n (w + \bar{z}_\mu - \bar{z})' m_\mu \right. \\ \left. - \frac{1}{2} \sum_{\mu=1}^n m_\mu' m_\mu \right\} dw$$

$$= C_0 \int_{\mathbb{R}^p} \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (w + \bar{z})' (w + \bar{z}) - \sum_{\mu=1}^n (w + \bar{z})' (z_\mu - \bar{z}) \right. \\ \left. - \frac{1}{2} \sum_{\mu=1}^n (z_\mu - \bar{z})' (z_\mu - \bar{z}) + \sum_{\mu=1}^n (w + \bar{z})' m_\mu \right. \\ \left. + \sum_{\mu=1}^n (z_\mu - \bar{z})' m_\mu - \frac{1}{2} \sum_{\mu=1}^n m_\mu' m_\mu \right\} dw$$

$$\text{or } \sum_{\mu=1}^n (w + \bar{z})' (z_\mu - \bar{z}) = (w + \bar{z})' \left[ \sum_{\mu=1}^n (z_\mu - \bar{z}) \right] = 0$$

$$f(z_2, \dots, z_n) = C_0 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (z_\mu - \bar{z})' (z_\mu - \bar{z}) + \sum_{\mu=1}^n (z_\mu - \bar{z})' m_\mu - \frac{1}{2} \sum_{\mu=1}^n m_\mu' m_\mu \right\} \\ \int_{\mathbb{R}^p} \exp \left\{ -\frac{n}{2} (w + \bar{z})' (w + \bar{z}) + \sum_{\mu=1}^n (w + \bar{z})' m_\mu \right\} dw$$

notons  $h(w)$  l'exponentielle située sous le signe intégral. Nous nous proposons de calculer

$$I_w = \int_{\mathbb{R}^p} h(w) dw$$

$$I_w = \prod_{i=1}^p \int_{\mathbb{R}} \exp \left\{ -\frac{n}{2} (w_i + \bar{z}_i)^2 + \sum_{\mu=1}^n (w_i + \bar{z}_i) m_{\mu i} \right\} dw_i$$

posons 
$$\omega_i = \sqrt{n} (w_i + \bar{z}_i) - \frac{1}{\sqrt{n}} \sum_{\mu=1}^n m_{\mu i}$$

$$d\omega_i = \sqrt{n} dw_i$$

$$I_w = \prod_{i=1}^p \frac{1}{\sqrt{n}} \left\{ \exp \frac{1}{2n} \left( \sum_{\mu=1}^n m_{\mu i} \right)^2 \right\} \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2} \omega_i^2 \right\} d\omega_i$$

$$= \left( \frac{2\pi}{n} \right)^{p/2} \exp \left\{ \frac{1}{2n} \sum_{i=1}^p \left( \sum_{\mu=1}^n m_{\mu i} \right)^2 \right\}$$

en définitive

$$f(z_1, \dots, z_n) = c_1 \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (z_\mu - \bar{z})' (z_\mu - \bar{z}) + \sum_{\mu=1}^n (z_\mu - \bar{z})' m_\mu \right. \\ \left. - \frac{1}{2} \sum_{\mu=1}^n m_\mu' m_\mu + \frac{1}{2n} \sum_{i=1}^p \left( \sum_{\mu=1}^n m_{\mu i} \right)^2 \right\}$$

posons 
$$B = (z_{n-p+1}, \dots, z_n)$$

alors B est presque sûrement inversible (voir remarque II.33)°

posons 
$$y_\mu = B^{-1} z_\mu \quad \text{pour } \mu = 2, \dots, n-p$$

Nous nous proposons de calculer la densité conjointe des  $n-p-1$  variables  $Y_2, \dots, Y_{n-p}$ . Nous utiliserons le résultat suivant :

#### IV.6. Lemme.

Soit  $Z_1 \dots Z_m$ ,  $m$  vecteurs de  $\mathbb{R}^p$ . A une matrice  $p \times p$  inversible  
et  $y_1 \dots y_m$ , vecteurs de  $\mathbb{R}^p$  tels que

$$Z_i = Ay_i \quad \text{pour } i = 1 \dots m$$

le jacobien de cette transformation est  $(\det A)^m$ .

(Preuve : le jacobien de la transformation pour  $i$  fixé

$$z_i = Ay_i \quad \text{est}$$

$$J = \frac{D(Ay_i)^1, \dots, (Ay_i)^P}{D(y_i^1, \dots, y_i^P)} = A$$

où  $(Ay_i)^j$  désigne la  $j^{\circ}$  composante du vecteur  $Ay_i \in \mathbb{R}^P$

Nous allons maintenant supposer que :

$$\begin{aligned} \frac{D(Ay_1, \dots, Ay_{m-1})}{D(y_1, \dots, y_{m-1})} &= \frac{D(Ay_1)^1, \dots, (Ay_1)^P, \dots, (Ay_{m-1})^1, \dots, (Ay_{m-1})^P}{D(y_1^1, \dots, y_1^P, \dots, y_{m-1}^1, \dots, y_{m-1}^P)} \\ &= (\det A)^{m-1} \end{aligned}$$

alors

$$\begin{aligned} \frac{D(Ay_1, \dots, Ay_m)}{D(y_1, \dots, y_m)} &= \det \begin{pmatrix} \boxed{A} & & & 0 \\ & \boxed{A} & & \\ & & \ddots & \\ & & & \boxed{A} \\ 0 & & & & \boxed{A} \end{pmatrix} \\ &= \det \begin{pmatrix} \boxed{A} & 0 & & 0 \\ & \ddots & & \\ 0 & & \boxed{A} & \\ \hline 0 & & & \boxed{I_P} \end{pmatrix} \times \begin{pmatrix} \boxed{I_P} & 0 & & 0 \\ & \ddots & & \\ 0 & & \boxed{I_P} & \\ \hline 0 & & & \boxed{A} \end{pmatrix} \\ &= \frac{D(Ay_1, \dots, Ay_{m-1})}{D(y_1, \dots, y_{m-1})} \times \det A = (\det A)^m \end{aligned}$$

c.q.f.d.)

Nous procédons comme précédemment : nous posons pour  $\mu = 2, \dots, n-p$

$$\begin{aligned} z_\mu &= By_\mu \quad \text{et pour } \mu = 1 \dots p \\ z_{n-p+\mu} &= V_\mu \end{aligned}$$

D'après le résultat bien connu sur la mesure image pour avoir la densité conjointe de  $Y_2, \dots, Y_{n-p}$ , on intègre alors par rapport à  $(v_1 \dots v_p) = B$

Remarquons que pour  $\mu = 1 \dots p$   $B^{-1}(z_{n-p+\mu}) = e_\mu$  où  $e_\mu$  désigne le  $\mu^{\circ}$  vecteur de base de  $\mathbb{R}^p$  ;

posons donc pour  $\mu = 1 \dots p$

$$y_{n-p+\mu} = e_{\mu}$$

et 
$$\bar{y} = \frac{1}{n} \sum_{\mu=1}^n y_{\mu} .$$

Opérons la transformation :

$$\begin{aligned} z_{\mu} &= B y_{\mu} \quad \text{pour } \mu = 1 \dots n \\ \bar{z} &= B \bar{y} \end{aligned}$$

D'après le lemme IV.6 le jacobien de cette transformation vaut en valeur absolue  $|\det B|^{n-p-1}$ ; nous noterons  $h(y)$ , la densité conjointe des  $n-p-1$  variables

$y_2 \dots y_{n-p}$

$$\begin{aligned} h(y) = C_1 \int_{\mathbb{R}^{p^2}} \exp \left\{ \frac{1}{2} \sum_{\mu=1}^n (y_{\mu} - \bar{y})' B' B (y_{\mu} - \bar{y}) \right. \\ \left. + \sum_{\mu=1}^n (y_{\mu} - \bar{y})' B' m_{\mu} + \frac{1}{2} \sum_{\mu=1}^n m_{\mu}' m_{\mu} + \frac{1}{2} \sum_{i=1}^p \left( \sum_{\mu=1}^n m_{\mu i} \right)^2 \right\} \\ |\det B|^{n-p-1} \quad dB \end{aligned}$$

d'où le résultat. )

#### IV.7. Corollaire.

Sous l'hypothèse  $H_{\sigma}$  la densité de l'invariant maximal  $h_{\sigma}(y)$  sous le groupe  $G_3$  nous est donnée par

$$\begin{aligned} h_{\sigma}(y) = C_1 \int \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (y_{\mu} - \bar{y})' B' B (y_{\mu} - \bar{y}) - c_2 \|\Delta m\|^2 \right. \\ \left. + \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (y_{\mu} - \bar{y})' B' \Delta m \right\} |\det B|^{n-p-1} \quad dB \end{aligned}$$

où

$$c_2 = \frac{\sum_{\mu=1}^n \alpha_{\sigma(\mu)}^2 - \left( \sum_{\mu=1}^n \alpha_{\sigma(\mu)} \right)^2}{2n}$$

(Preuve : l'invariant maximal sous le groupe  $G_3$  nous est justement donné par  $y = y_2, \dots, y_{n-p}$  (voir proposition II.32.) Sous l'hypothèse  $H_{\sigma}$ , on a

$$m_{\mu} = \alpha_{\sigma(\mu)} \Delta m \quad \text{pour } \mu = 1 \dots n$$

On en conclut le corollaire.)

Définissons maintenant  $u_1, \dots, u_p$  comme étant les  $p$  vecteurs colonnes tels que

$$B' = (u_1, \dots, u_p)$$

IV.8. Lemme.

$$\sum_{\mu=1}^n (y_{\mu} - \bar{y})' B' B (y_{\mu} - \bar{y}) = \sum_{i=1}^p u_i' \mathcal{Y} u_i$$

où  $\mathcal{Y}$  est la matrice  $p \times p$  symétrique non singulière p.s :

$$\mathcal{Y} = \sum_{\mu=1}^n (y_{\mu} - \bar{y}) (y_{\mu} - \bar{y})'$$

(Preuve :

$$\sum_{\mu=1}^n (y_{\mu} - \bar{y})' B' B (y_{\mu} - \bar{y}) = \sum_{\mu=1}^n \sum_{h,k=1}^p a_{hk} (y_{\mu h} - \bar{y}_h) \cdot (y_{\mu k} - \bar{y}_k)$$

où  $a_{hk}$  est le terme situé à l'intersection de la ligne  $h$  et de la colonne  $k$  de la matrice symétrique  $B'B$ . Intervertissons les signes  $\sum$ . Il vient

$$\sum_{h,k} a_{hk} \sum_{\mu=1}^n (y_{\mu h} - \bar{y}_h) (y_{\mu k} - \bar{y}_k),$$

or  $a_{hk} = \sum_{j=1}^p u_{jh} u_{jk}$  ; on obtient alors

$$\sum_{h,k} \left( \sum_{j=1}^p u_{jh} u_{jk} \right) \sum_{\mu=1}^n (y_{\mu h} - \bar{y}_h) (y_{\mu k} - \bar{y}_k)$$

$$= \sum_{j=1}^p \sum_{hk} u_{jh} u_{jk} \mathcal{Y}_{hk} \quad \text{avec}$$

$$\mathcal{Y}_{hk} = \sum_{\mu=1}^n (y_{\mu h} - \bar{y}_h) (y_{\mu k} - \bar{y}_k)$$

$$= \sum_{j=1}^p t_{uj} \mathcal{Y}_{uj} \quad . )$$

Soit  $\mathcal{Y}^{1/2}$  telle que  $\mathcal{Y}^{1/2} \mathcal{Y}^{1/2} = \mathcal{Y}$

opérons alors la transformation :

$$r_j = \mathcal{Y}^{1/2} u_j \iff u_j = \mathcal{Y}^{-1/2} r_j \quad du_j = |\det \mathcal{Y}|^{-1/2} dr_j$$

pour  $j = 1 \dots p$

$$h_{\sigma}(y_2 \dots y_{n-p}) = C_1 \left[ \exp \left\{ -\frac{1}{2} \sum_{j=1}^p t_{u_j} u_j - c_2 \|\Delta m\|^2 \right. \right. \\ \left. \left. + \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (y_{\mu} - \bar{y}) B' \Delta m \right| \det B \right|^{n-p-1} dB$$

posons  $\mathcal{C}_{\sigma} = \mathcal{Y}^{1/2} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (y_{\mu} - \bar{y}) \right]$

et soit D telle que

$$D = B \mathcal{Y}^{1/2}$$

alors on a  $D' = \mathcal{Y}^{1/2} \quad B' = \mathcal{Y}^{1/2} (u_1, \dots, u_p) = (\mathcal{Y}^{1/2} u_1, \dots, \mathcal{Y}^{1/2} u_p) = (r_1, \dots, r_p)$

$$B' = \mathcal{Y}^{-1/2} D'$$

$$B = D \mathcal{Y}^{-1/2}$$

$$\det B = \det D \det \mathcal{Y}^{-1/2}$$

$$\det \mathcal{Y}^{1/2} = (\det \mathcal{Y})^{-1/2}$$

#### IV.9. Lemme III.3.

On a les identités suivantes

$$1^{\circ} \sum_{j=1}^p u'_j \mathcal{Y} u_j = \sum_{j=1}^p r'_j r_j$$

$$2^{\circ} \mathcal{C}'_{\sigma} \mathcal{C}_{\sigma} = \|\mathcal{C}_{\sigma}\|^2 = \frac{T_{\sigma}^2}{n-1}$$

(Preuve :

$$1^{\circ} \sum_{j=1}^p u'_j \mathcal{Y} u_j = \sum_{j=1}^p u'_j \mathcal{Y}^{1/2} \mathcal{Y}^{1/2} u_j \\ = \sum_{j=1}^p r'_j r_j$$

$$\begin{aligned} 2^\circ \mathcal{C}'_\sigma \mathcal{C}_\sigma &= \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (y_\mu - \bar{y}) \right]' \mathcal{J}^{-1} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (y_\mu - \bar{y}) \right] \\ &= \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (z_\mu - \bar{z}) \right]' (B^{-1})' \mathcal{J}^{-1} B \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (z_\mu - \bar{z}) \right] \end{aligned}$$

or

$$\begin{aligned} \mathcal{Y} &= \sum_{\mu=1}^p (y_\mu - \bar{y}) (y_\mu - \bar{y})' = \sum_{\mu=1}^n B^{-1} (z_\mu - \bar{z}) (z_\mu - \bar{z})' (B^{-1}) \\ &= B^{-1} \sum_{\mu=1}^n (z_\mu - \bar{z}) (z_\mu - \bar{z})' (B^{-1})' \\ &= B^{-1} \sum_{\mu=1}^n (x_\mu - \bar{x}) {}^t(x_\mu - \bar{x}) (B^{-1})' \end{aligned}$$

d'où  $\mathcal{J}^{-1} = \frac{1}{n-1} B' S^{-1} B$

d'où  $\mathcal{C}'_\sigma \mathcal{C}_\sigma = \frac{1}{(n-1)} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_\mu - \bar{x}) \right]' S^{-1} \left[ \sum_{\mu=1}^n \alpha_{\sigma(\mu)} (x_\mu - \bar{x}) \right] = \frac{T_\sigma^2}{n-1}$

$$\mathcal{C}'_\sigma \mathcal{C}_\sigma = \frac{T_\sigma^2}{n-1} \quad .)$$

L'expression de la densité devient :

$$\begin{aligned} h_\sigma(y_2 \dots y_{n-p}, \Delta m) &= c_1 |\det \mathcal{J}|^{-\frac{n-1}{2}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^p r_j' r_j \right. \\ &\quad \left. - c_2 \|\Delta m\|^2 + \mathcal{C}'_\sigma D' \Delta m \right\} |\det D|^{n-p-1} d D \end{aligned}$$

IV.10. Remarque.

Sous l'hypothèse  $H_b$  la densité s'écrit

$$h_b(\mathcal{Y}) = c_1 |\det \mathcal{J}|^{-\frac{n-1}{2}} \exp \left\{ -1/2 \sum_{j=1}^p {}^t r_j r_j \right\} |\det D|^{n-p}$$

en effet : sous l'hypothèse  $H_0 \quad \Delta m = 0$

IV.11. Lemma.

La fonction  $h_\sigma(y_2, \dots, y_{n-p}, \Delta m)$  définie ci-dessus est une fonction

croissante de  $T_\sigma^2$

(Preuve :

$$h_{\sigma}(\underline{y}) = C_1 |\det \mathcal{P}|^{-\frac{n-1}{2}} \exp \{ -c_2 \|\Delta m\|^2 \} \times$$

$$\int \exp \left\{ -\frac{1}{2} \sum_{j=1}^p r'_j r_j + \mathcal{C}'_{\sigma} D' \Delta m \right\} |\det D|^{n-p-1} dD$$

effectuons une notation  $P$  de telle sorte que

$$P \mathcal{C}'_{\sigma} = \|\mathcal{C}'_{\sigma}\| e_1 = \frac{1}{n-1} T_{\sigma} e_1$$

et

$$t_j = Pr_j$$

et définissons la matrice

$$\Gamma = PD' = (t_1, \dots, t_p)$$

où les  $t_i$  sont des vecteurs colonnes pour  $i = 1 \dots p$  alors

$$h_{\sigma}(\underline{y}, \Delta m) = c_1 |\det \mathcal{P}|^{-\frac{n-1}{2}} \exp \{ -c_2 \|\Delta m\|^2 \} \times$$

$$\int \exp \left\{ -\frac{1}{2} \sum_{j=1}^p t'_j \cdot t_j + \frac{1}{n-1} T_{\sigma} e_1' \Gamma \Delta m \right\} |\det \Gamma|^{n-p-1} d\Gamma$$

posons

$$J(\sigma, \Delta m) = \int \exp \left\{ -\frac{1}{2} \sum_{j=1}^p t'_j \cdot t_j + \frac{1}{n-1} T_{\sigma} e_1' \Gamma \Delta m \right\} |\det \Gamma|^{n-p-1} d\Gamma$$

on va montrer que  $J(\sigma, \Delta m)$  est fonction croissante de  $T_{\sigma}^2$  : écrivons l'exponentielle sous la forme

$$\exp \left\{ -\frac{1}{2} \sum_{j=1}^p t'_j \cdot t_j \right\} \exp \left\{ \frac{1}{n-1} T_{\sigma} \langle e_1, \Gamma \Delta m \rangle \right\}$$

$$= \exp \left\{ -\frac{1}{2} \sum_{j=1}^p t'_j \cdot t_j \right\} \exp \left\{ \frac{1}{n-1} T_{\sigma} \times (\Gamma \Delta m)_1 \right\}$$

où  $(\Gamma \Delta m)$  est la 1<sup>o</sup> composante du vecteur  $\Gamma \Delta m$

$$= \exp \left\{ -\frac{1}{2} \sum_{j=1}^p t'_j \cdot t_j \right\} \left( \sum_{\ell=0}^{+\infty} \frac{\left(\frac{1}{n-1} T_{\sigma}\right)^{\ell}}{\ell!} (\Gamma \Delta m)_1^{\ell} \right)$$

ceci en écrivant l'exponentielle sous forme de série.

Intégrons, il vient

$$J(\sigma, \Delta m) = \sum_{\ell=0}^{+\infty} C_{\ell} \left(\frac{1}{n-1} T_{\sigma}\right)^{\ell}$$

où

$$C_{\ell} = \frac{1}{\ell!} \int \exp \left\{ -\frac{1}{2} \sum_{j=1}^p t'_j \cdot t_j \right\} (\Gamma \Delta m)_1^{\ell} |\det \Gamma|^{n-p-1} d\Gamma$$

Montrons que les coefficients  $C_{2\ell+1}$  sont nuls, effectuons le changement de variable

$$\Psi = \begin{pmatrix} -1 & & 0 \\ & \ddots & \\ & & -1 \\ 0 & & & -1 \end{pmatrix} \Gamma = (\psi_1, \dots, \psi_p)$$

où les  $\psi_j$   $j = 1 \dots p$  sont  $p$  vecteurs de  $\mathbb{R}^p$

alors

$$C_{2\ell+1} = \frac{1}{2^{+1}!} \int \exp \left\{ -\frac{1}{2} \sum_{j=1}^p \psi'_j \psi_j \right\} (-\psi \Delta m)_1^{2\ell+1} |\det \Psi|^{n-p-1} d\Psi$$

$$= - C_{2\ell+1}$$

d'où  $C_{2\ell+1} = 0$

le fait que  $J(\sigma, \Delta m)$  est une fonction croissante de  $T_\sigma^2$  est à présent établi :

$$J(\sigma, \Delta m) = \sum_{\ell=0}^{+\infty} C_{2\ell} \left( \frac{1}{n-1} T_\sigma \right)^{2\ell} = \sum_{\ell=0}^{+\infty} C_{2\ell} (T_\sigma^2)^\ell \left( \frac{1}{n-1} \right)^{2\ell}$$

Il en résulte que  $h_\sigma(y)$  est une fonction croissante de  $T_\sigma^2$  .)

Nous définissons sur  $\mathcal{X}$  la probabilité a priori  $b$  telle que

$$b[H_0] = b_0$$

$$\text{et } b[H_\sigma] = \frac{1 - b_0}{n!}$$

Soit  $c \in (\mathbb{R}_*^+)$ , nous définissons la mesure  $\eta_c(\Delta m)$  comme étant la mesure uniforme sur la sphère de  $\mathbb{R}^p$   $\|\Delta m\|^2 = c$ . Nous prenons comme fonction de coût la fonction  $L_0$  définie chapitre II §2 et recherchons les règles doubles de Bayes invariantes sous  $G_3$  relatives à la probabilité a-priori  $b \otimes \eta_c$  sur  $\mathcal{X} \times \mathbb{R}^p$ .

En utilisant la caractérisation II.41, nous savons que la règle double de Bayes invariante sous  $G_3$ , unique à une équivalence près, nous est donnée par

$$\begin{aligned}
 (IV.12.) \quad d_0(x) &= 1 \quad \text{si } Ah_0(y) > \int h_{\sigma}(y, \Delta m) d\eta_c(\Delta m) \quad \text{pour tout } \sigma \in \mathfrak{G} \\
 d_{\sigma}(x) &= 1 \quad \text{si } \int h_{\sigma}(y, \Delta m) d\eta_c(\Delta m) > Ah_0(y) \\
 &\quad \text{et si } \int h_{\sigma}(y, \Delta m) d\eta_c(\Delta m) > \int h_{\sigma'}(y, \Delta m) d\eta_c(\Delta m) \\
 &\quad \text{pour tout } \sigma' \in \mathfrak{G}, \sigma' \neq \sigma
 \end{aligned}$$

où A est une constante de  $\mathbb{R}_*^+$  qui dépend de a le niveau de signification du

test. or

$$\begin{aligned}
 &\int_{\mathbb{R}^p} h_{\sigma}(y, \Delta m) d\eta_c(\Delta m) \\
 &= C_1 |\det \mathcal{Y}|^{-\frac{n-1}{2}} e^{-c \cdot c_2} \int_{||\Delta m||^2=c} \exp \left\{ -\frac{1}{2} \sum_{j=1}^p r_j r_j + \mathcal{C}'_{\sigma} D' \Delta m \right\} \\
 &\quad |\det D|^{n-p-1} dD d\eta_c(\Delta m)
 \end{aligned}$$

D'après le lemme IV.11. et compte tenu de l'expression des densités  $h_0$  et  $h_{\sigma}(y, \Delta m)$  on conclut comme en III.3. les règles doubles de Bayes relatives à la probabilité a-priori  $b \otimes \eta_c$  sur  $\mathcal{X} \times \mathbb{R}^p$ , invariants sous  $G_3$  sont caractérisés par

$$\begin{aligned}
 d_0(x) &= 1 \quad \text{si } T_{\sigma}^2 = \max_{\sigma \in \mathfrak{G}} T_{\sigma}^2 < \lambda_a \\
 d_{\sigma}(x) &= 1 \quad \text{si } T_{(\sigma)}^2 > \lambda_a
 \end{aligned}$$

où  $\lambda_a$  est une constante qui dépend du niveau de signification du test a.

Cette règle de décision est maintenant symétrique en  $x_1, \dots, x_n$ . On en conclut que d'après la symétrie des densités  $f_{\sigma}(x_1 \dots x_n)$  des lois conjointes  $X_1, \dots, X_n$  sous l'hypothèse  $H_{\sigma}$  que la condition 4° est vérifiée.

Enfin en utilisant le lemme III.8. avec  $\Lambda = \int_p$  on conclut que les probabilités de décision  $P(D_{\sigma}/d, (H_{\sigma}, \Delta m))$  associée à la règle ci-dessus ne dépendant que de  $\Delta m$ . En appliquant alors le théorème II.20., puis en faisant varier  $c \in \mathbb{R}_*^+$  on obtient le théorème IV.3.)

## §2 APPLICATIONS

Comme dans le cas  $\Lambda$  connue, nous donnons ici des corollaires du théorème IV qui nous permettent l'un de retrouver un résultat connu, les autres de généraliser à plusieurs dimensions des résultats établis en dimension 1.

### IV.13. Corollaire.

Sous l'hypothèse qu'il n'y a au plus qu'une valeur aberrante, dans le cadre du modèle  $\mathcal{H}$  décrit ci-dessus, la règle de décision satisfaisant aux hypothèses 1-2-3-4 et possédant les propriétés du théorème IV nous est donnée par

prendre la décision  $D_0$  si  $T^2_{(n)} < \lambda_a$

prendre la décision  $D_{(n)}$  si  $T^2_{(n)} > \lambda_a$

où 
$$T^2_{(n)} = \max_{\mu=1 \dots n} (x_\mu - \bar{x})' S^{-1} (x_\mu - \bar{x})$$

Nous retrouvons ici un résultat connu et établi par Karlin, S. et Truax, D.R. (1960).

### IV.14. Corollaire.

Supposons que  $1 < k < n/2$  des observations proviennent d'un même glissement de la valeur moyenne  $\Delta m \in \mathbb{R}^p$ . Alors la règle de décision satisfaisant aux hypothèses du théorème IV et possédant les propriétés qui y sont énoncées, nous est donnée par

prendre la décision  $D_0$  si  $T^2_{(\sigma)} < \lambda_a$

prendre la décision  $D_{(\sigma)}$  si  $T^2_{(\sigma)} > \lambda_a$

où 
$$T^2_{(\sigma)} = \max \left[ \sum_{\mu=1}^k (x_{\sigma(\mu)} - \bar{x}) \right]' S^{-1} \left[ \sum_{\mu=1}^k (x_{\sigma(\mu)} - \bar{x}) \right]$$

On obtient ici la généralisation en dimension  $p$  du résultat bien connu de Murphy, R.B. (1951) en dimension 1.

Supposons que les  $n$  observations soient partagées en  $r$  classes de  $k$  éléments. Nous nous proposons de savoir si l'un des classes a glissé en moyenne.

IV.15. Corollaire.

La règle de décision correspondante, vérifiant les hypothèses du théorème IV et possédant les propriétés qui y sont énoncées, nous est donnée par

prendre la décision  $D_0$  si  $T^2_{(\ell)} < \lambda'_a$

prendre la décision  $D_{(\ell)}$  si  $T^2_{(\ell)} > \lambda'_a$

où  $\lambda'_a = \lambda_a/k^2$  et  $(\ell)$  désigne l'indice pour lequel

$$T^2_{\ell} = [\bar{x}(\ell) - \bar{x}]' S^{-1} [\bar{x}(\ell) - \bar{x}]$$

est maximal,  $1 \leq \ell \leq r$ ,  $\bar{x}(\ell)$  étant la moyenne de la  $\ell^{\circ}$  classe.

La démonstration de ce corollaire se calque sur celle du corollaire III.11. On obtient ici la généralisation en dimension  $p$  du résultat bien connu de Paulson, E. (1952) en dimension 1 .

§3 GLISSEMENTS ORTHOGONAUX

Comme au chapitre III §3 nous envisageons maintenant le cas de deux valeurs aberrantes éventuelles provenant dans le cadre du modèle  $\mathcal{A}$  de glissements orthogonaux. Nous retrouvons l'analogie du théorème III.14. en remplaçant  $\Lambda$  par  $S$  la matrice estimée sans biais de  $\Lambda$  . Nous précisons ici brièvement la formulation du problème et les hypothèses que nous faisons.

Soient  $X_1 \dots X_n$   $n$  variables aléatoires gaussiennes à valeurs dans  $\mathbb{R}^p$  de moyennes respectives  $m_1 \dots m_n$  et de même matrice des variances-covariances supposées inconnues. Nous voulons tester l'hypothèse  $H_0 = H(m_1 = \dots = m_n = m)$  contre les hypothèses alternatives que nous noterons  $H_{\nu\xi}$  que les valeurs  $x_{\nu}$  et  $x_{\xi}$  ont glissé en moyenne de deux vecteurs de  $\mathbb{R}^p$  :  $\Delta m_1$  et  $\Delta m_2$  respectivement, ces vecteurs étant inconnus mais orthogonaux au sens du produit scalaire associé à la matrice  $\Lambda^{-1}$  . Le problème est donc symétrique en  $\Delta m_1$  et  $\Delta m_2$  . En conséquence nous supposons  $1 \leq \nu < \xi \leq n$  . Nous avons donc  $\frac{n(n-1)}{2}$  hypothèses alternatives  $H_{\nu\xi}$  .

$$H_{\nu\xi} = H(m_1 = \dots = m_{\nu} - \Delta m_1 = \dots = m_{\xi} - \Delta m_2 = \dots = m_n = m \quad \nu < \xi )$$

Nous formulons les hypothèses 1-2-3- et 4 du §1 en respectant nos notations nouvelles. Sous ces hypothèses nous noterons  $h_{\nu\xi}(y, \Delta m_1, \Delta m_2)$  la densité de l'invariant maximal sous le groupe  $G_3$  sous l'hypothèse  $H_{\nu\xi}$ . Nous appellerons  $D_{\nu\xi}$  la décision d'agir comme si  $H_{\nu\xi}$  était vraie.

Un règle de décision  $d$  nous est alors donnée par

$$d = (d_0, d_{1,2}, \dots, d_{n-1,n})$$

avec

$$d_0 + \sum_{\substack{\nu < \xi \\ \nu, \xi = 1}}^n d_{\nu\xi}(x_1 \dots x_n) = 1$$

les  $d_{\nu\xi}$  étant des fonctions mesurables par rapport à la mesure de Lebesgue sur  $\mathbb{R}^{p \cdot n}$ . Nous avons alors le résultat suivant :

IV.16. Théorème.

Dans le cadre du modèle  $\mathcal{K}$ , la règle de décision  $d$  satisfaisant aux hypothèses 1-2-3 et 4 du §1 qui maximise les probabilités de décision

$P(D_{\nu\xi}/d, (H_{\nu\xi}; \Delta m_1, \Delta m_2))$  uniformément en  $\nu, \xi$  et  $\Delta m_1, \Delta m_2$  nous est donnée par

$$d(x) = 1 \quad \text{si} \quad T_{(n)}^2 + T_{(n-1)}^2 < \lambda_a$$

$$d_{(n)(n-1)}(x) = 1 \quad \text{si} \quad T_{(n)}^2 + T_{(n-1)}^2 > \lambda_a$$

ou  $\lambda_a$  est une constante satisfaisant à l'hypothèse 1,  $(n)(n-1)$  sont les deux indices pour lesquels on obtient les deux plus grandes valeurs de

$$T_{\mu}^2 = (x_{\mu} - \bar{x})' S^{-1} (x_{\mu} - \bar{x}) .$$

(Preuve : la performance des règles de décision invariantes sous  $G_3$  est indépendante de  $m$  et de  $\Lambda$ . Nous supposons donc  $m = 0$ ,  $\Lambda = \mathbb{1}_p$  .

Nous allons identifier l'espace des paramètres  $(\Delta m_1, \Delta m_2)$  à  $\mathbb{R}^p \oplus \mathbb{R}^p$ . Pour cela posons  $\Delta m_s = \Delta m_1 \oplus \Delta m_2$ . Le produit scalaire canonique lié à cette somme directe nous est associé à la matrice bloc

$$\begin{pmatrix} \mathbb{1}_p & 0 \\ 0 & \mathbb{1}_p \end{pmatrix} = \mathbb{1}_{2p}$$

Sur les vecteurs  $\Delta m_s$  nous définissons pour  $c \in \mathbb{R}_*^+$  fixé la probabilité a priori

$\eta_c(\Delta m_s)$  concentrée sur l'ellipsoïde  $\otimes ||\Delta m_s||^2 = c$ . Sur  $\mathcal{H}$  nous définissons la probabilité a priori  $b$  telle que

$$b[H_0] = b_0$$

$$b[H_{v\xi}] = \frac{2(1-b_0)}{n(n-1)}$$

Nous allons rechercher les règles doubles de Bayes relatives à la probabilité a priori  $b \otimes \eta_c$  définie sur  $\mathcal{H} \times \mathbb{R}^{2p}$ , invariantes sous  $G_3$ .

Pour cela nous avons besoin de calculer la densité de l'invariant maximal sous  $G_3$  sous les diverses hypothèses  $H_{v\xi}$ .

IV.17. Lemme.

Sous l'hypothèse  $H_{v\xi}$ , la densité de l'invariant maximal sous le groupe  $G_3$  vaut

$$h_{v\xi}(Y, \Delta m_s) = c_1 \exp \left\{ -\frac{n-1}{2n} \otimes ||\Delta m_s||^2 \right.$$

$$\left. \exp \left\{ -\frac{1}{2} \sum_{\mu=1}^n (y_\mu - \bar{y})' B' B (y_\mu - \bar{y}) + \otimes \langle (y_v - \bar{y}), \begin{pmatrix} B' & 0 \\ 0 & B \end{pmatrix} \Delta m_1 \otimes \Delta m_2 \rangle \right\} \right.$$

$$\left. | \det B |^{n-p-1} dB \otimes \langle (Y_v - \bar{y}) \otimes (Y - \bar{y}), \rangle \right.$$

selon les notations du lemme IV.5.

(Preuve : à partir de l'expression du lemme IV.5. on pose

$$m_v = \Delta m_1, m_\xi = \Delta m_2, m_\mu = 0 \quad \mu = 1 \dots n \quad v \neq \mu \neq \xi$$

En tenant compte de  $\langle \Delta m_1, \Delta m_2 \rangle = 0$  on trouve facilement le résultat.)

Nous allons quelque peu modifier l'expression de  $h_{v\xi}$  en tenant compte en particulier des lemmes IV.8. - IV.9. et des notations introduites.

Posons de plus

$$\mathcal{L}_{v\xi} = \begin{pmatrix} \mathcal{J}^{-1/2} & 0 \\ 0 & \mathcal{J}^{-1/2} \end{pmatrix} [(y_v - \bar{y}) \otimes (y_\xi - \bar{y})]$$

Remarquons que  $\otimes ||\mathcal{L}_{v\xi}||^2 = \frac{1}{n-1} (T_v^2 + T_\xi^2)$ . Alors en posant

$$D'_b = \begin{pmatrix} D' & 0 \\ 0 & D' \end{pmatrix}$$

il vient

$$h_{v\xi}(y, \Delta m_s) = C_1 |\det \mathcal{J}|^{-\frac{n-1}{2}} \exp \left\{ -\frac{n-1}{2} \bullet \|\Delta m_s\|^2 \right\}$$

$$\int \exp \left\{ -\frac{1}{2} \sum_{j=1}^p r'_j r_j + \mathcal{C}_{v\xi} D'_b \Delta m_s \right\} |\det D|^{n-p-1} dD$$

de la même façon qu'en IV.11. nous démontrons le résultat suivant.

IV.18. Lemme.

La fonction  $h_{v\xi}(y, \Delta m_s)$  ci-dessus est une fonction croissante de  $\|\mathcal{C}_{v\xi}\|^2$

La densité  $h_0(y)$ , quant à elle, nous est fournie par la remarque IV.10.

La caractérisation II.41. nous fournit la règle double de Bayes :

$$d_0(x) = 1 \quad \text{si } Ah_0 > \int h_{v\xi} d\eta_C(\Delta m_s) \quad \forall v, \xi = 1 \dots n$$

$$d_{v\xi}(x) = 1 \quad \text{si } \int h_{v\xi} d\eta_C(\Delta m_s) > Ah_0$$

$$\int h_{v\xi} d\eta_C(\Delta m_s) > \int h_{\mu\zeta} d\eta_C(\Delta m_s)$$

$$\forall \mu \neq v \quad \zeta \neq \xi \quad \mu < \zeta \quad v < \xi \quad \text{et } \mu, v, \xi, \zeta = 1 \dots n$$

$$\int h_{v\xi}(y, \Delta m_s) d\eta_C(\Delta m_s) = C_1 \exp \left\{ -\frac{n-1}{2n} c \right\} \int_{\bullet \|\Delta m_s\|^2 = c} \exp \left\{ -\frac{1}{2} \sum_{j=1}^p r'_j r_j + \mathcal{C}_{v\xi} D'_B \Delta m_s \right\} |\det D|^{n-p-1} dD d\eta_C(\Delta m_s)$$

On conclut alors comme en III.3. que la règle double de Bayes relative à la probabilité a priori  $b \bullet \eta_C$ , invariante sous  $G_3$  nous est fournie par

$$d_0(x) = 1 \quad \text{si } T_{(n)}^2 + T_{(n-1)}^2 < \lambda_a$$

$$d_{(n)(n-1)}(x) = 1 \quad \text{si } T_{(n)}^2 + T_{(n-1)}^2 > \lambda_a$$

Cette règle est manifestement symétrique en  $x_1 \dots x_n$ . D'après la symétrie des densités des lois conjointes des variables aléatoires  $X_1 \dots X_n$  sous les hypothèses  $H_{v\xi}$  la condition 4 est vérifiée.

Enfin comme au lemme III.8. on montre que  $P(d_{v\xi}/d, (H_{v\xi}, \Delta m_s))$  ne dépend que de

•  $\|\Delta m_s\|$  . En appliquant alors le théorème II.20. puis en faisant varier  $C$  on obtient le théorème IV.16.).

#### §4 CONCLUSION

Nous venons d'obtenir un certain nombre de résultats lorsque cette fois la matrice  $\Lambda$  est inconnue. Ces résultats sont en tout similaire à ceux du chapitre III si ce n'est par le remplacement de  $\Lambda^{-1}$  par  $S^{-1}$  . Le résultat du paragraphe 3 s'étend facilement au cas où il a  $k$  valeurs aberrantes éventuelles, avec  $k > 2$  , qui proviennent de glissements orthogonaux deux à deux.

Cependant en l'absence de toute information sur les glissements on peut dire qu'il n'y a pas de règle optimale de détection.



## CHAPITRE V

### §0 INTRODUCTION

Nous voulons étudier ici les problèmes auxquels doit faire face concrètement le statisticien lorsqu'il désire effectuer une détection de valeurs aberrantes éventuelles dans un échantillon. Nous discutons : du traitement d'un échantillon multidimensionnel composante par composante, de l'hypothèse de normalité ; nous décrivons le traitement de deux échantillons ; enfin nous suggérons une solution pour le problème lié à une application répétée de la règle de Thompson généralisée évoqué en I.4.

### §1 TRAITEMENT D'UN ECHANTILLON MULTIDIMENSIONNEL

#### COMPOSANTE PAR COMPOSANTE

Pour mieux fixer les idées nous prenons  $p = 2$  et supposons avoir affaire à une loi  $\mathcal{N}_2(0, I_2)$ . La densité du couple  $(x_{\mu_1}, x_{\mu_2})$ ,  $\mu^\circ$  vecteur de l'échantillon est donc égale à  $\frac{1}{2\pi} e^{-\frac{1}{2}(x_{\mu_1}^2 + x_{\mu_2}^2)}$ .

Nous voulons comparer ici une application globale de l'analogue de la règle de Thompson généralisée lorsque  $\Lambda$  est connue, avec une application composante de cette règle. Nous nous situons dans le cas théorique  $n = 1$ . Etudier si cette valeur est aberrante ou non revient à regarder si le point  $(x_1, x_2)$  est à l'intérieur d'un cercle (C) tel que  $P[(X_1, X_2) \in (C)] = 1 - a$  où  $a$  est le niveau de signification.

En ajustant bien sûr les niveaux de signification, traiter le problème composante par composante revient à regarder si cette valeur est à l'intérieur du carré  $[-R, +R] \times [-R, +R]$  où  $R$  est le rayon du cercle (C).

La différence entre un traitement composante par composante et un traitement global provient que dans le premier cas on ne détecte pas les valeurs situées à l'intérieur de la zone hachurée: Z.H. Pour une loi normale, en choisissant  $R$

tel que  $P[(X_1, X_2) \in (C)] = 95 \% \longrightarrow R = 2,4477$

on a  $P[(X_1, X_2) \in Z.H.] = 2,1 \%$

C'est-à-dire qu'en testant un grand nombre de valeurs, en admettant qu'elles vérifient toutes une loi  $\mathcal{N}_2(O, I_2)$  on ne détectera pas, dans 2,1 % des cas, des valeurs qui devraient être détectées. Mais justement, une valeur aberrante ne vérifie pas la même loi que la loi de base. Elle vérifie par exemple, dans le cas de glissement, une loi  $\mathcal{N}_2(\Delta m, I_2)$ .

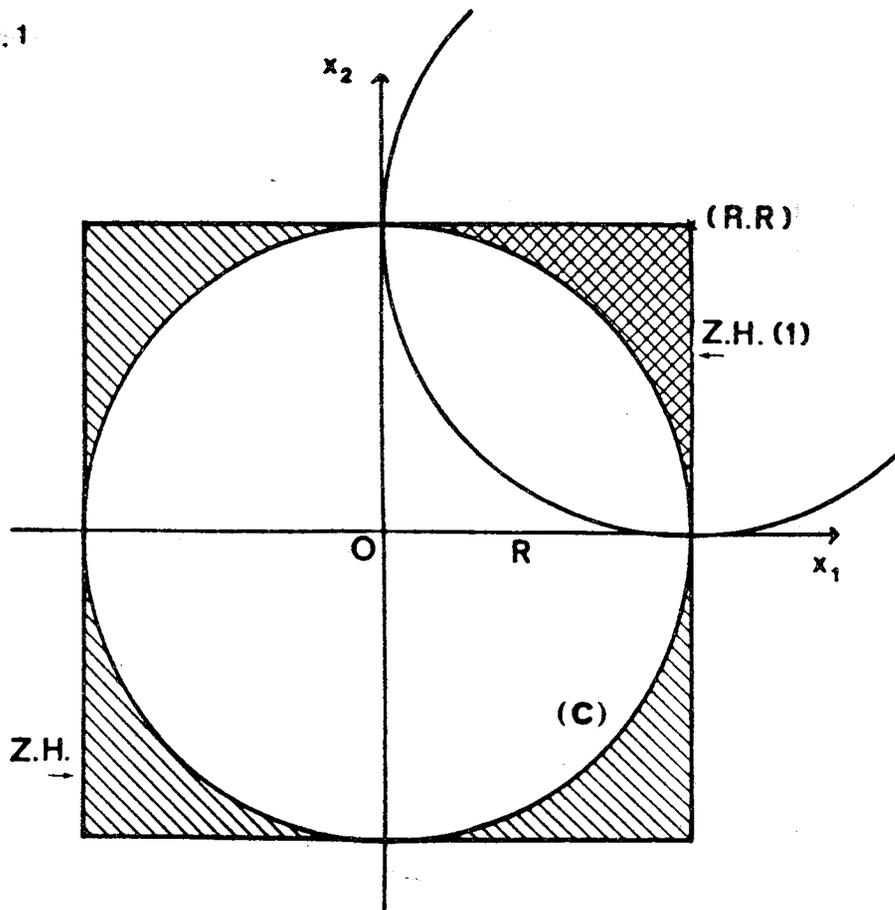
Supposons  $\Delta m = (R, R)$ , où  $R$  est calculé ci-dessus. La probabilité qu'une valeur aberrante se situe dans  $Z.H.(1) = P[(Y_1, Y_2) \in Z.H.(1)]$

où  $(Y_1, Y_2) \sim \mathcal{N}_2([R, R], I_2) \neq 0,19$

Nous considérons  $P[(Y_1, Y_2) \in Z.H. - Z.H.(1)]$  comme négligeable.

Si l'on teste un grand nombre d'échantillons dans lequel se trouvent  $l$  valeurs aberrantes de loi  $\mathcal{N}_2((R, R), I_2)$  en moyenne dans 19 % des cas on ne détectera pas cette valeur. Cet exemple illustre bien les limites d'un traitement multidimensionnel composante par composante.

DESSIN V.1



## §2 TESTS DE NORMALITE

Nous ne disposons pas d'outils pour tester la normalité globale d'un échantillon multidimensionnel. Nous testons donc cette hypothèse dans certaines directions et l'admettons globalement. La puissance de nos tests est basée sur l'hypothèse de normalité. Il faudrait donc en étudier la tolérance lorsque l'on s'écarte de cette hypothèse.

En dimension 1 nous disposons par contre d'outils opérationnels (ex :  $\chi^2$ , droite de Henry). Cependant la présence de valeurs aberrantes dans l'échantillon peut rendre ces tests négatifs. Il convient donc de les pratiquer avant et après le tri. L'idéal est de disposer d'un échantillon dont on est sûr que la loi de probabilité associée est de Laplace - Gauss ; ceci grâce à des études préalables. C'est rarement le cas.

## §3 ECHANTILLON I.N.S.E.E.

Nous avons pris contact avec l'Unité de recherche de l'I.N.S.E.E. Nous étudions des échantillons multidimensionnels log-normaux, dont l'un est décrit ci-dessous avec le traitement correspondant. Les travaux ne sont pas encore terminés.

Cet échantillon est constitué à partir de 252 entreprises dont on connaît trois paramètres : travail, capital, production sur 11 ans de 1963 à 1973.

\* Ces données ont été fournies sur la bande magnétique n° 1 403 par l'I.N.S.E.E. le 10.X.75. Ne disposant d'aucune indication sur les entreprises, nous les avons numérotées de 1 à 252 dans l'ordre selon lequel les données correspondantes ont été enregistrées.

Le traitement a été effectué en dimension 3 année par année. On a supposé, sans le vérifier, que ces onze échantillons de taille 252 suivaient une loi log-normale de  $\mathbb{R}^3$ , dont on a estimé la moyenne et la matrice des variances-covariances à l'aide de l'échantillon. Nous avons alors calculé pour chacune des entreprises l'écart généralisé studentisé  $T^2$  à la valeur moyenne, c'est-à-dire que si l'on

\* voir page 110

appelle  $\bar{x}$  la moyenne estimée,  $S$  la matrice des variances-covariances estimée sans biais et  $x_v$  le vecteur de  $\mathbb{R}^3$  dont les composantes sont dans l'ordre le logarithme du travail, le logarithme du capital et le logarithme de la production, de la  $v^0$  entreprise, on a la relation

$$T^2 = {}^t(x_v - \bar{x}) S^{-1} (x_v - \bar{x})$$

Nous avons ensuite appliqué la règle de Thompson généralisée, en choisissant un niveau de confiance  $\alpha = 0,01$  auquel correspond la valeur critique 10.93.

La règle de Thompson généralisée, consiste à considérer comme aberrant tout vecteur  $x_v$  tel que  $T_v^2 > 10,93$ .

Le tableau V.2. nous fournit année par année les entreprises dont l'écart généralisé studentisé  $T_v^2$  dépasse la valeur critique.

Le listing ci-joint résume les différents résultats année par année. Sont fournis dans l'ordre :

- les composantes du vecteur moyenne  $\bar{x}$
- la matrice des variances-covariances estimée sans biais  $S$
- le déterminant  $DS$  de  $S$
- l'inverse  $S^{-1}$  de la matrice  $S$
- les écarts généralisés studentisés  $T_v^2$  en ordre croissant, ainsi que le numéro des entreprises correspondantes
- la somme des écarts généralisés studentisés.

Cette somme est une constante dont la valeur théorique vaut  $(n-1)p$ ,  $n$  étant la taille de l'échantillon, et,  $p$  sa dimension. Ici somme =  $251 \times 3 = 753$ .

Cette somme a été calculée pour servir de vérification à nos calculs. Le tableau II nous donne la différence entre cette valeur théorique et celle effectivement donnée par notre programme, en fonction de l'année.

Cette différence est toujours inférieure à 0,9, la somme calculée approchant toujours par défaut la somme théorique. En admettant que la valeur de l'écart généralisé studentisé calculée, approche par défaut la valeur réelle, l'erreur moyenne

par défaut par entreprise est inférieure à  $1/252 < 4 \times 10^{-3}$  par défaut ce qui paraît suffisant en tenant compte des résultats figurant sur le listing, et du fait que la valeur critique ne nous est fournie qu'avec 2 chiffres significatifs après la virgule.

Une première remarque nous est fournie par l'entreprise 132. En effet l'écart généralisé studentisé  $T_{1,32}^2$  de l'année 1965 vaut 163,37. En consultant les données, nous constatons que le logarithme de la production de 1965 de cette entreprise a été remplacé par 0.0. Le logarithme du travail étant de 3,91. En consultant les valeurs du même paramètre de l'entreprise 132 de l'année précédente, on voit que le logarithme du travail de 1964 de l'entreprise vaut 4,00, celui de la production 8,37. Cette donnée est donc fantaisiste.

A cause d'un effet de Cache évident, il convient

- soit de refaire un passage machine avec la donnée exacte
- soit de supprimer cette entreprise de façon à effectuer une détection plus sûre.

En consultant le tableau V.2. nous trouvons quatre types de détection :

- 1° Des entreprises rejetées successivement année par année. Il s'agit des entreprises 93, 196, 225. Ces entreprises ont certainement des caractéristiques nettement différentes du reste et ne doivent sans doute pas être retenues pour des études statistiques à partir de l'échantillon.
- 2° Nous trouvons ensuite les entreprises 26 et 44 qui sont détectées jusqu'en 1968 et 1969 respectivement puis "rentrent dans le rang".
- 3° Par contre un troisième groupe d'entreprises (4, 228, 240) ne sont détectées qu'à partir de 1966 pour la première et de 1969 pour les deux autres, et semblent avoir subi une évolution contraire au groupe précédent : il n'est pas encore possible de dire si cette évolution autour des années charnières 67-69 est suffisamment nette ou si au contraire les groupes 2 et 3 seront détectés, chaque année lors d'un 2° passage.

4° Enfin, un quatrième groupe d'entreprises ne sont détectées qu'épisodiquement. Il s'agit des entreprises 95, 110, 130, 132 , 136, 184, 201, 202, 204, 226, 233, 243.

Il serait intéressant d'effectuer un second passage machine en ôtant les diverses entreprises détectées année par année ou en ôtant systématiquement toutes les entreprises détectées au moins une fois sur les onze années d'étude, de façon à donner une certaine homogénéité aux données restantes.

On pourra appliquer certaines méthodes pour assurer la convergence de notre procédure séquentielle - pour le calcul des estimateurs en particulier - Nous pensons par exemple à une winsorisation généralisée.

On pourra également procéder à des tests de log-normalité sur l'échantillon ainsi trié.

\* Nous reproduisons ici le compte-rendu qui avait été adressé à l'I.N.S.E.E.

CARACTERISTIQUES DU TRI

Application de la règle de Thompson généralisée  
11 fois (de 1963 à 1973) - 1° passage machine

Taille de l'échantillon :  $n = 252$   
Dimension de l'échantillon :  $p = 3$   
Niveau de signification choisi :  $\alpha = 0,01$   
Valeur critique correspondante :  $\lambda_{0,01} = 10,93$   
Indice numérotant les entreprises :  $v \quad v = 1, \dots, 252$   
Nature des vecteurs :  $x_v$

$$x_v = \left( \begin{array}{l} \text{logarithme du travail} \\ \text{logarithme du capital} \\ \text{logarithme de la production} \end{array} \right) \text{ de la } v^{\circ}$$

entreprise pour une année (de 1963 à 1973)

Hypothèse de base :  $x_v$  est une réalisation d'une loi log-normale  
à valeurs sans  $\mathbb{R}^3$ .

Estimé du vecteur moyenne  $\bar{x} = \frac{1}{252} \sum_{v=1}^{252} x_v$

Estimé sans biais de la matrice des variances-covariances

$$S = \frac{1}{251} \sum_{v=1}^{252} (x_v - \bar{x}) {}^t(x_v - \bar{x})$$

Ecart généralisé studentisé à la valeur moyenne de la  $v^{\circ}$  entreprise

$$T_v^2 = {}^t(x_v - \bar{x}) S^{-1} (x_v - \bar{x})$$

Somme des écarts  $\sum_{v=1}^{252} T_v^2 = (n-1)p = 753$

TABLEAU V.2.

V A L A B II 2

p = 3    n = 252     $\lambda$  0,01 = 10,93

Années	n° des entreprises détectées (entreprises pour lesquelles $T_v^2 > 10,93$ )										Années		
	26	202	196	233	225	93							
1963	26	202	196	233	225	93							
1964	"		"	"	"	"	44						
1965			"		"	"		110	132				
1966	"		"		"	"	"	"		4			
1967	"	"	"		"	"	"	"		"	226	243	136
1968	"		"		"	"	"	"		"	"	204	
1969			"		"	"	"	"			240	228	
1970			"		"	"	"	"		"	"	95	
1971			"		"	"	"	"		"	"	184	
1972			"		"	"	"	"		"	"	"	130
1973			"		"	"	"	"		"	"	"	201

N.B. - Le n° des entreprises apparaît l'année de leur première détection  
 Pour indiquer une détection lors d'une année ultérieure nous plaçons le symbole " dans la case correspondante.

TABLEAU V.3.

Années	Somme des écarts généralisés studentizés $\sum_{v=1}^{252} T_v^2$	Différence entre $(n-1)p$ et $\sum_{v=1}^{252} T_v^2$
1963	752,76	0,24
1964	752,38	0,62
1965	752,75	0,25
1966	752,80	0,20
1967	752,91	0,09
1968	752,12	0,88
1969	752,19	0,81
1970	752,21	0,79
1971	752,56	0,44
1972	752,76	0,24
1973	752,76	0,24
maximum des différences 0,88		

#### §4 DEUXIEME ECHANTILLON

L'un des principaux échantillons que nous avons étudiés nous a été fourni par les résultats du premier tour des élections présidentielles de 1974, en ne conservant que deux des candidats : MM. GISCARD d'ESTAING et MITTERAND. Les données sont, par départements, les différences entre le pourcentage départemental et le pourcentage national des voix de chaque candidat, arrondis à deux chiffres après la virgule. Nous y avons ajouté les données correspondantes des départements et territoires d'Outre-Mer. Les données sont présentées au tableau V.4. Nous avons testé la normalité composante par composante et suivant la première bissectrice, à l'aide de la droite de Henry. Les fréquences cumulées sont fournies par le tableau V.5. et les diagrammes correspondants par les tableaux V.6, 7, 8. Nous avons appliqué la règle de Thompson généralisée, avec un niveau de confiance  $1 - \alpha = 95 \%$ . L'application brute et répétée de ce test est décrite en I.4.

Une telle application nous prive de beaucoup d'information.

Que se passe-t-il exactement ?

A la suite d'une première détection, pour procéder à une nouvelle application, nous avons besoin d'estimer une moyenne et une matrice des variances-covariances. Nous le faisons à l'aide de l'échantillon privé d'un certain nombre de valeurs. Or en dimension  $p = 1$ , Dixon, W.J. (1960) a montré que de tels estimés sont peu efficaces, moins en ce qui concerne la moyenne, qu'en ce qui concerne l'écart type.

Des palliatifs ont été proposés par lui, également par Tukey, J.W. (1960) "contribution to discussion of papers by f.e. Anscombe et Daniel" *Technometrics* 2. p. 160-165 - qu'ils ont appelés winsorisation et qui reviennent à effectuer un certain coefficient aux valeurs détectées de façon à diminuer leur poids lors d'une estimation. Les calculs de Dixon montrent qu'alors, les estimés obtenus par cette méthode sont efficaces.

En dimension  $p > 1$  nous proposons une méthode que nous appelons winsorisation généralisée et qui consiste à remplacer une valeur détectée, par la valeur qui lui est la plus proche au sens de la distance généralisée. Ceci revient en fait, à

DIFFERENCES ENTRE POURCENTAGE DEPARTMENTAL ET POURCENTAGE NATIONAL

tableau V 4

DISCAP D ESTANG	MITFRANK	DEPARTEMENTS MMEPOTES	
4.27	-3.41	ATN	1
-9.70	5.21	ATNF	2
0.66	5.72	ALLIEN	3
-2.47	4.82	ALPES-DE-HAUTE-PROVENCE	4
-0.40	-0.74	ALPES-MAITRES	5
6.44	0.74	ALPES-MARITIMES	6
0.12	-1.15	ARDENNES	7
-1.22	4.71	ARDENNES	8
-14.23	13.40	ARIEGE	9
1.44	-0.25	AISNE	10
-11.54	13.02	AISNE	11
8.00	-5.70	AVOYON	12
-2.72	8.42	BOUCHES-DU-RHONE	13
4.47	-4.85	CALVADOS	14
13.00	-10.76	CANTAL	15
-10.04	1.15	CHARENTE	16
-7.05	-1.44	CHARENTE-MARITIME	17
-1.40	1.70	CHER	18
-4.71	6.40	CHARENTE	19
-8.13	1.71	CORSE	20
3.38	-0.13	COTE D'OR	21
2.00	1.03	COTE-DU-NORD	22
-3.14	6.40	CREUSE	23
-12.05	4.43	DORDOGNE	24
5.19	-1.40	DROME	25
-2.55	2.57	DROME	26
1.05	-0.43	FINDE	27
2.53	-2.22	FINDE-ET-LIRA	28
8.05	-6.14	FINIETOF	29
-3.57	7.42	GARD	30
-3.02	6.45	GARONNE (HAUTE)	31
-9.50	7.54	GERE	32
-18.04	-0.87	GIRONNE	33
-7.40	6.50	HAUTE-LOIRE	34
10.00	-0.00	HAUTE-LOIRE	35
-6.05	0.70	INDRE	36
-12.72	-8.14	INDRE-ET-LOIRE	37
-1.74	7.00	ISERE	38
2.00	-0.54	JURA	39
-17.00	1.41	LANDES	40
-1.75	-2.75	LOIRE-ET-CHER	41
3.71	-1.70	LOIRE	42
11.11	-8.74	LOIRE (HAUTE)	43
5.47	-6.24	LOIRE-ATLANTIQUE	44
3.33	-5.10	LOIRET	45
-9.12	3.82	LOT	46
-10.22	3.04	LOT-ET-GARONNE	47
14.72	-11.44	LOZERE	48
7.43	-12.47	MAINE-ET-LOIRE	49
11.04	-13.70	MARNE	50
1.05	-0.05	MARNE	51
1.34	-1.74	MARNE (HAUTE)	52
12.17	-14.40	MAYENNE	53
4.05	3.20	MELANIE-ET-MOSELLE	54
4.05	-1.47	MELANIE	55
13.70	-0.53	MORBIHAN	56
6.00	-0.82	MORBIHAN	57
-6.20	13.77	MORBIHAN	58
-3.70	4.81	MORBIHAN	59
-3.23	2.15	NICE	60
5.00	-10.12	NICE	61
-7.24	8.77	PAS-DE-CALAIS	62
12.44	-0.75	PAYS-DE-LOIRE	63
-6.54	-6.14	PYRENEES-ATLANTIQUES	64
-8.50	7.40	PYRENEES (HAUTES)	65
-3.00	8.05	PYRENEES-ORIENTALES	66
11.00	-12.21	DUIN (OAS)	67
7.74	-14.10	DUIN (HAUTE)	68
4.81	-1.24	DUIN	69
2.82	1.77	SARRE (HAUTE)	70
0.21	2.31	SARRE-ET-LOIRE	71
-1.47	-1.83	SARTHE	72
-1.75	0.12	SAVOIE	73
8.42	-7.75	SAVOIE (HAUTE)	74
4.04	-5.01	SARTE	75
-0.04	3.03	SEINE-MARITIME	76
-0.78	-0.42	SEINE-ET-MARNE	77
4.24	-2.42	SEINE	78
2.10	-8.70	SEVRES (DEUX)	79
-5.41	4.04	SEVRES	80
-3.25	1.03	TADU	81
-6.57	1.48	TADU-ET-GARONNE	82
1.97	2.80	VAR	83
-2.77	4.80	VANNES	84
9.43	-15.40	VANNES	85
-7.01	-3.70	VANNES	86
-8.54	10.20	VANNES (HAUTE)	87
4.00	-3.17	VANNES	88
3.12	-1.57	VANNES	89
-2.14	4.28	VANNES DE BELFORT	90
-1.02	3.43	VANNES	91
2.11	0.18	VAL-DE-SEINE	92
-7.78	11.73	SEINE-SAINT-DENIS	93
-2.52	5.77	VAL-DE-MARNE	94
-1.78	4.40	VAL-D'oise	95
-14.77	7.00	GUANDELONNE	96
-15.25	-0.74	GUANDELONNE	97
-18.77	-8.74	MARTINIQUE	98
-16.00	4.53	REUNION	99
-23.48	-18.40	COMORES	100
-23.55	-25.55	AFRICA ET ISSAS	101
-7.44	-0.72	MELANIE-CALDONNIE-MERIDIONALES	102
-11.58	7.00	PRIYANESIE FRANCAISE	103
5.38	-21.79	SAINT PIERRE ET MOUTFLOU	104
7.15	-41.54	WALLIS ET FUTUNA	105

TABLEAU V.5.

Ce tableau donne les fréquences cumulées en fonction des écarts entre les % départementaux et le % national des voix de MM. GISCARD et MITTERAND au 1<sup>o</sup> tour des élections présidentielles de 1974 ; ainsi que les fréquences cumulées en fonction de la somme de ces écarts.

%	Fréquences cumulées × 100		
	GISCARD	MITTERAND	Somme
- 14	7,6	6,7	
- 13	8,6	7,6	
- 12	10,5	9,5	
- 11	12,4	10,5	
- 10	14,3	12,4	10,5
- 9	15,2	14,3	10,5
- 8	18,1	18,1	15,2
- 7	21,9	19,0	16,2
- 6	23,8	21,9	21
- 5	26,7	24,8	21
- 4	29,5	25,7	26,7
- 3	38,1	28,6	27,6
- 2	43,8	31,4	30,5
- 1	51,4	40,0	40
0	54,3	51,4	45,7
1	57,1	55,2	50,5
2	61,9	62,9	63,8
3	67,6	66,7	75,2
4	71,4	72,4	88,6
5	76,2	81,0	92,4
6	80	83,8	94,3
7	84,8	87,6	97,1
8	87,6	92,4	99
9	90,5	95,2	
10	91,4	95,2	
11	92,4	96,2	
12	95,2	97,1	
13	97,1	97,1	
14	99	100	

V.6. Diagramme des écarts concernant M. GISCARD D'ESTAING

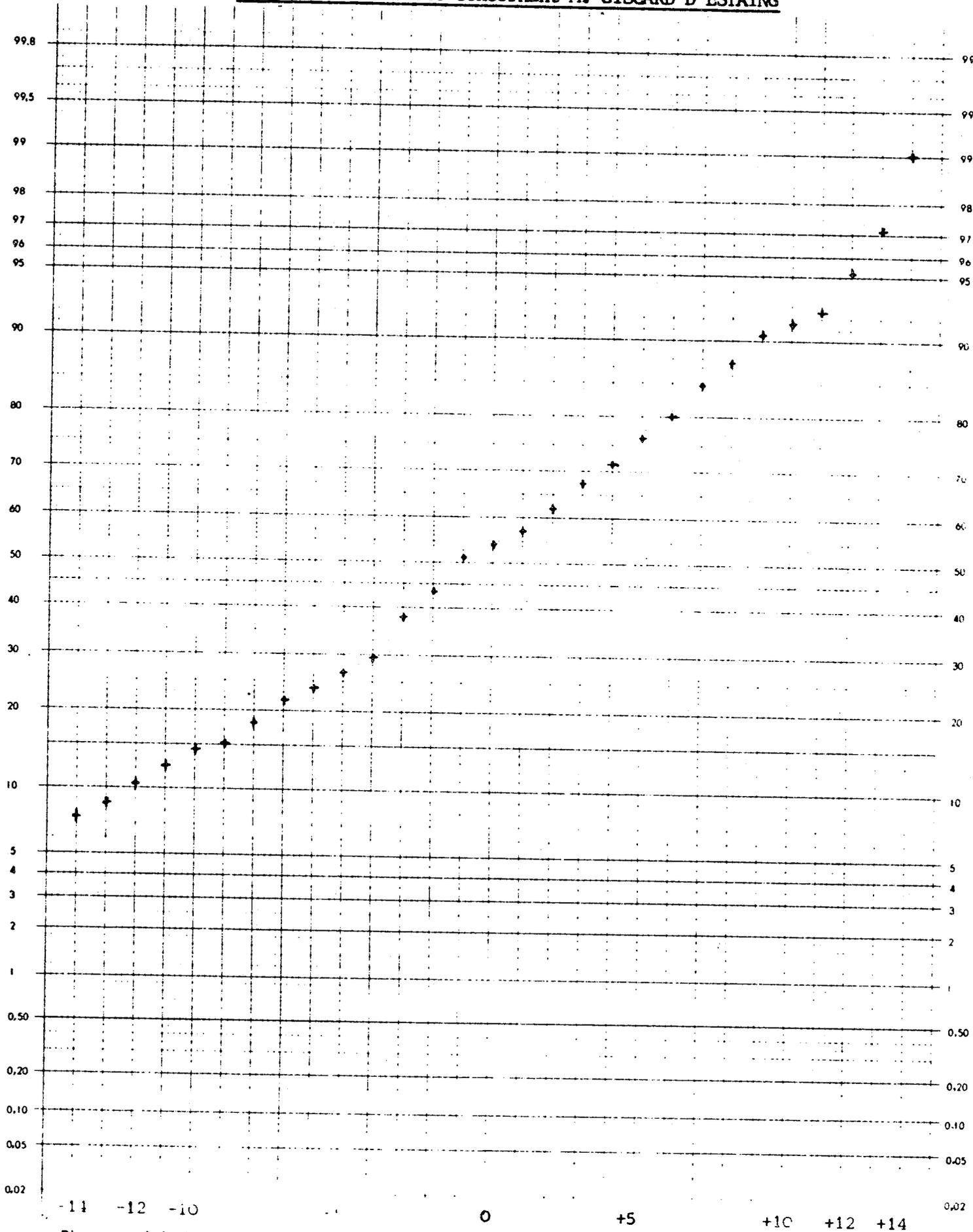


Diagramme de la droite de Henry ou droite-échantillon (anamorphe de la courbe de probabilité totale de la loi de Gauss)

V.7. Diagramme des écarts concernant M. MITTERAND

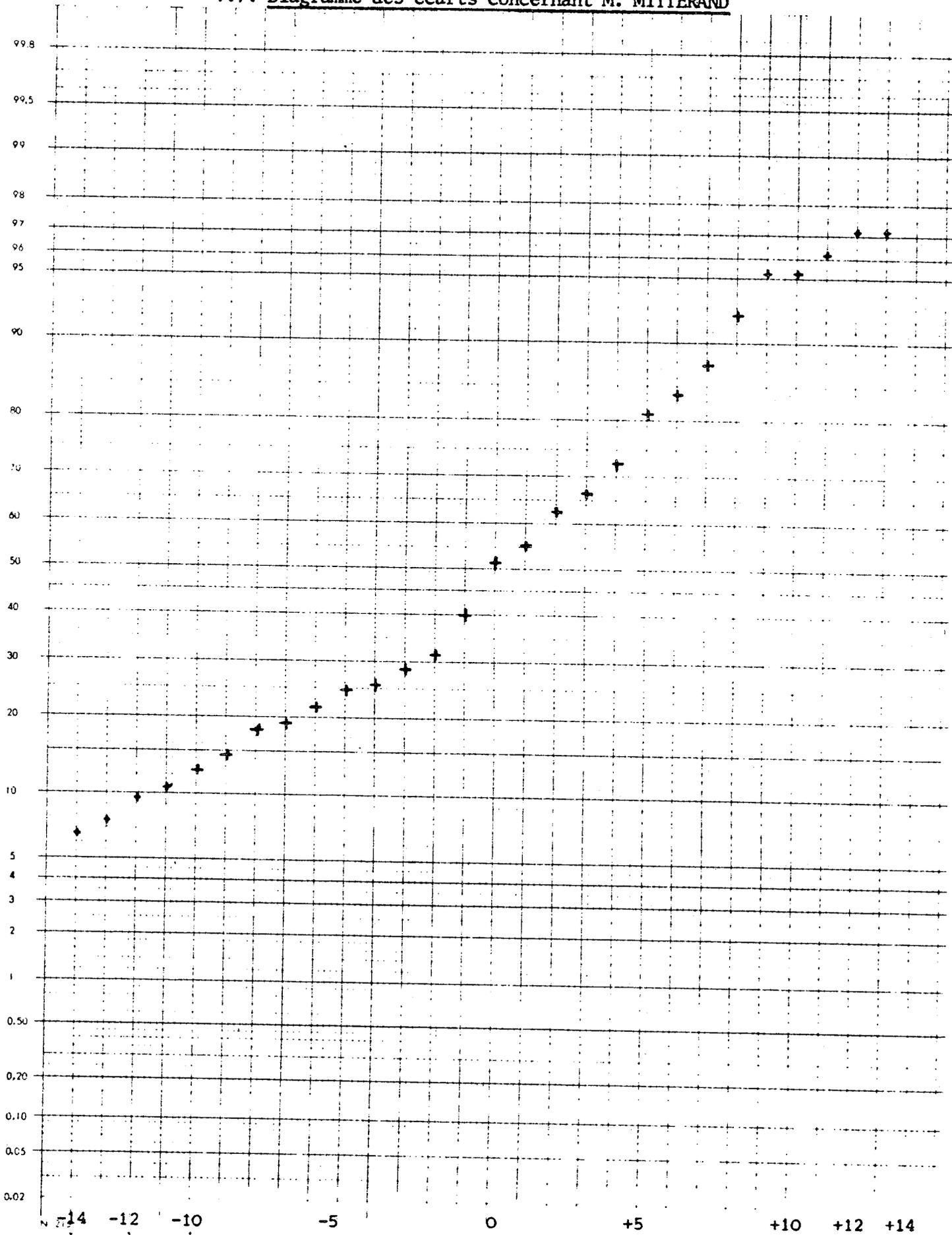


Diagramme de la droite de Henry ou droite-échantillon (anomorphe de la courbe de probabilité totale de la loi de Gauss.)

V.8. Diagramme de la somme des écarts

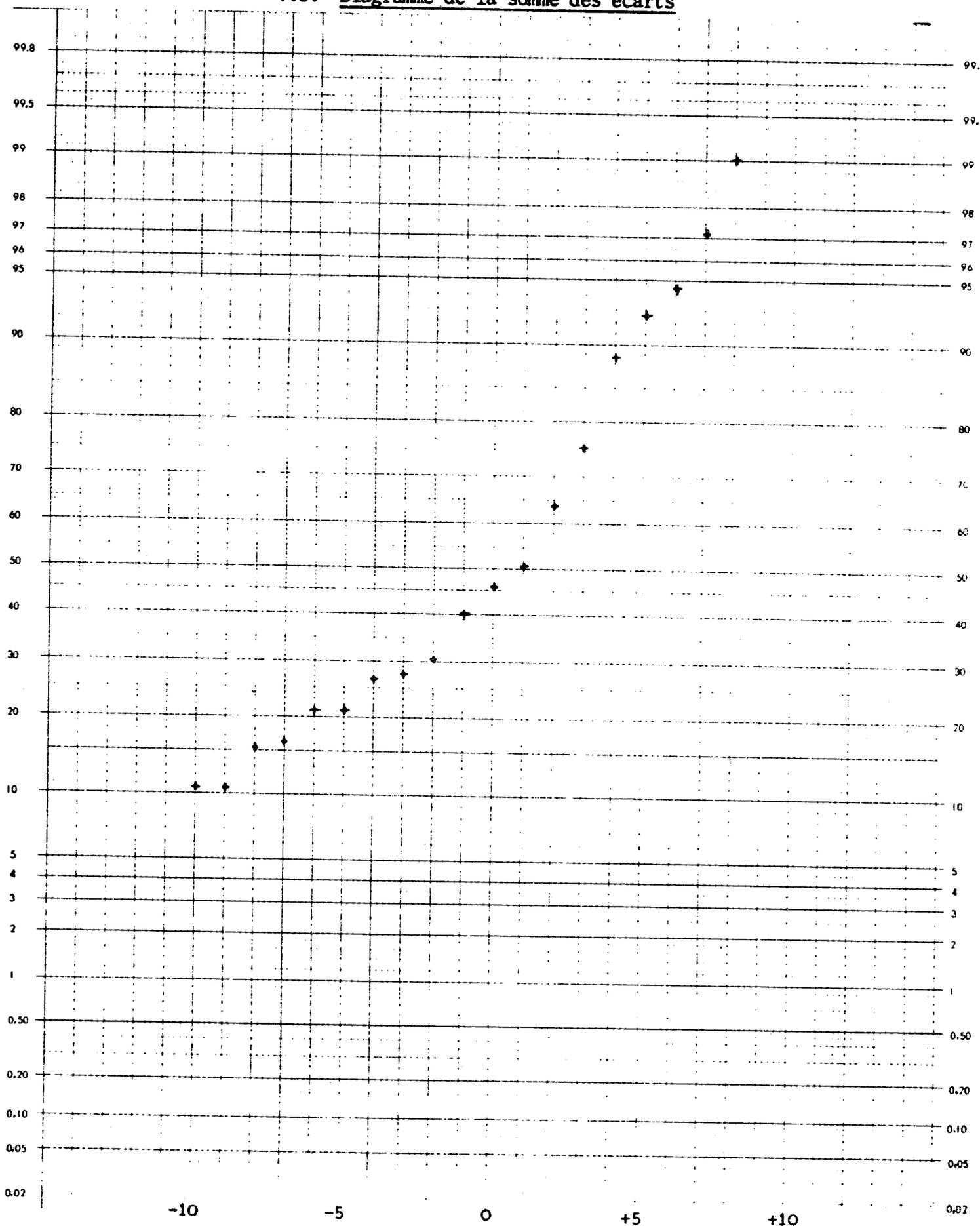
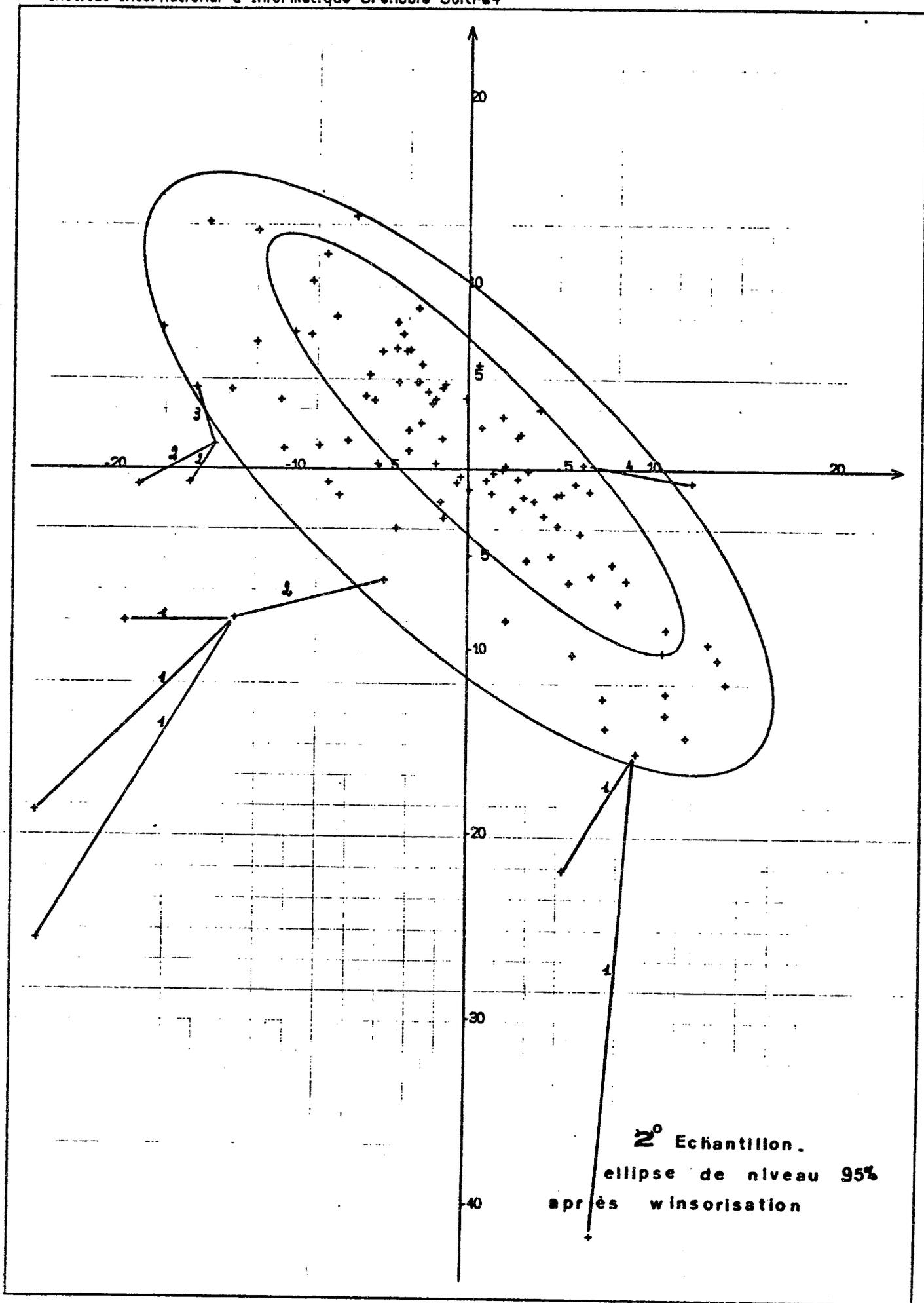


Diagramme de la droite de Henry ou droite-échantillon (anamorphe de la courbe de probabilité totale de la loi de Gauss).



affecter un certain coefficient à chacune des composantes de l'observation détectée ; ce coefficient étant fonction de la composante et du nuage de points.

Cette méthode a pour avantage, lors d'une application répétée de la règle de Thompson généralisée :

- a) de conserver une taille fixe à l'échantillon
- b) de ne nécessiter le calcul que d'une valeur critique ; d'où un gain de place et de temps machine
- c) de stopper plus rapidement la méthode.

Sur notre échantillon, cette méthode fait stopper la détection au bout de 4 passages après avoir détecté 10 observations, nous en laissant plus de 90 %. Bien entendu, toute estimation ultérieure est faite à partir des observations non détectées et des valeurs fournies lors du dernier passage (sans détection) pour remplacer celles qui avaient été détectées.

L'algorithme de la règle de Thompson généralisée avec winsorisation est fourni en V.10.

V.10. Algorithme de la règle de Thompson généralisée avec winsorisation généralisée

$x_1 \dots x_n$  n-échantillon de loi  $\mathcal{N}_p(m, \Lambda)$

n taille de l'échantillon

p dimension de l'échantillon

m,  $\Lambda$  moyenne et matrice des variances-covariances de la loi associée à l'échantillon. m,  $\Lambda$  sont inconnus

a niveau de rejet

1 - Calcul de la valeur critique  $\lambda_a$

pour a choisi soit  $F_{p;n-p;a}$  la valeur critique de niveau a de la loi F à p et n - p degrés de liberté

$$\lambda_a = \frac{(n-1)^2 F_{p;n-p;a}}{\frac{n(n-p)(n-2)}{(n-1)p} + nF_{p;n-p;a}}$$

2 - Calcul du vecteur moyenne estimé pour  $i = 1 \dots p$

$$\bar{x}_i = \frac{1}{n} \sum_{\mu=1}^n x_{\mu i}$$

3 - Calcul de la matrice estimée sans biais des variances-covariances

$$S_{ij} = \frac{1}{n-1} \sum_{\mu=1}^n (x_{\mu i} - \bar{x}_i)(x_{\mu j} - \bar{x}_j)$$

4 - Calcul de la matrice inverse  $S^{-1}$  de la matrice des variances-covariances

5 - Calcul des écarts généralisés studentisés à la valeur moyenne  $\bar{x}$  pour  $\mu = 1 \dots n$

$$T_{\mu}^2 = (x_{\mu} - \bar{x})' S^{-1} (x_{\mu} - \bar{x})$$

6 - Calcul de la somme des écarts généralisés

$$T.S. = \sum_{\mu=1}^n T_{\mu}^2$$

(pour contrôle des calculs. Valeur théorique T.S. = (n-1)p)

7 - Rangements des écarts en ordre décroissant et sortie

8 - Détection des valeurs aberrantes et calcul du nombre N det de valeurs détectées :

pour  $\mu = 1 \dots n$ , si  $T_{\mu}^2 \geq \lambda_a$  N det := N det + 1

9 - Si nombre de valeur détectée = 0 aller en 12 sinon aller en 10

10 - Winsorisation. Pour toute observation  $x_{\mu_0}$  tel que  $T_{\mu_0}^2 \geq \lambda_a$ , calcul de la valeur la plus proche au sens de la distance généralisée, parmi les valeurs non détectées.

$$D_{\mu}^2(\mu_0) = (x_{\mu_0} - x_{\mu})' S^{-1} (x_{\mu_0} - x_{\mu})$$

pour  $\mu \neq \mu_0$  et  $\mu \neq$  des indices des valeurs détectées.

Recherche de  $D_{(1)}^2(\mu_0) = \min_{\mu} D_{\mu}^2$

11 - Préparation du passage suivant : pour toute valeur détectée  $x_{\mu_0}$ , remplacer  $x_{\mu_0}$  par  $x_{(1)}$  où (1) est l'indice pour lequel  $D_{\mu}^2(\mu_0)$  est minimal. Aller en 2

12 - fin.

On aura intérêt à utiliser des témoins indiquant le numéro du passage, la détection, le remplacement d'une observation par une autre, etc...

Faire les calculs en double précision.

## §5 SUGGESTIONS

Cette contribution nous montre qu'en ce qui concerne le traitement des valeurs extrêmes multidimensionnelles, beaucoup de travail reste à faire. Nous suggérons à titre d'exemple : une étude de la winsorisation généralisée ; la tabulation de certains tests ; la comparaison dans le cas de deux valeurs aberrantes éventuelles dans un échantillon, des trois tests suivants :

1° double application du test de Karlin et Truax-Ferguson

2° test de Wilks dans le cas de deux observations douteuses

3° test proposé en IV.14. lorsque  $\alpha_1 = \alpha_2 = 1$   $\alpha_\mu = 0$  pour  $\mu \neq 1,2$

Nous savons déjà que lorsque les valeurs aberrantes éventuelles proviennent d'un glissement en moyenne identique, le test 3 est optimal.



B I B L I O G R A P H I E

- ANDERSON, T.W. (1958). An introduction to multivariate statistical analysis. J. Wiley and Sons, New-York. 374 pages.
- ANSCOMBE, F.J. (1960). Rejection of outliers - *Technometrics* 2, pp.123-147
- BARRA, J.R. (1971). Notions fondamentales de statistique mathématique Dunod, Paris. 258 p.
- BIRNBAUM, A. (1959). On the analysis of factorial experiments without replication. *Technometrics* 1, pp.343-357.
- CHAMBADAL, L.  
OVAERT, J.L. (1968). Algèbre linéaire et algèbre tensorielle. Dunod, Paris. 544 p.
- CHAUVENET, W. (1863). *Manual of spherical and practical Astronomy*. Philadelphie. Appendice sur la méthode des moindres carrés. L'appendice fut réédité en 1868 sous le titre : A treatise on the method of least squares.
- DANIEL, C. (1960). Locating outliers in factorial experiments. *Technometrics* 2, pp. 149-156.
- DAVID, H.A. (1970) Order Statistics. J. Wiley and Sons, New-York. 272 p.
- DIXON, W.J. (1960). Simplified estimation from censored normal samples. *A.M.S.* 31, pp. 385-91.
- FERGUSON, T.S. (1961a). Rules for rejection of outliers. *Revue Inst. int. de Stat.* 29 pp.29-43.
- FERGUSON, T.S. (1961b). On the rejection of outliers. *Proc. 4th Berkeley Sump.* I, 253-287.
- FOURGEAUD, C. FUCHS, A. (1967). Statistique. Dunod, Paris. 325 p.
- GAREL, B.B. (1975). Valeurs aberrantes dans un échantillon gaussien multidimensionnel. Exposé Grenoble.
- GEFFROY, J. (1958-59). Contribution à la théorie des valeurs extrêmes. *Publications de l'I.S.U.P.* Fasc. 7 - pp. 37-121, Fasc 8, pp. 123-185.
- GNAHADESIKAN,  
R. KETTERING, J.R. (1972). Robust estimates, residuals, and outliers detection with multiresponse data. *Biometrics* 28, pp. 81-124.
- GRUBBS, F.E. (1950). Sample criteria for testing outlying observations. *A.M.S.* 21, pp. 27-58.

- GRUBBS, F.E.(1969). Procedure for detecting outlying observations in samples. *Technometrics*, vol. 11, n° 1. pp. 1-21.
- HAJEK, J.(1967). Theory of rank tests. Academic press, New-York, 297 p.
- HAWKINS, D.M.(1974) The detection of error in multivariate Data using principal components. *J. of Am. Stat. Association*. Vol. 69. n° 346, pp. 340-44.
- KARLIN, S. RUBIN, H.(1956). The theory of decision procedure for distributions with monotone likelihood ratio. *A.M.S.* 27, pp. 272-300.
- KARLIN, S. TRUAX, D.R.(1960). Slippage problems. *A.M.S.* 31, pp. 448-62.
- KRUSKALL, W.H.(1960). Some remarks on wild observations. *Technometrics*, vol. 2, n° 1. pp. 1-3.
- KUDO, A.(1956a). On the testing of outlying observations. *Sankhya*, 17, 67-76.
- KUDO, A.(1956b). On the invariant multiple decision procedures. *Bull. Math. Stat.* 6, pp. 57-68.
- KUDO, A.(1957). The extreme value in a multivariate normal sample. *Mem. Fac. Sci. Kyushu Univ.* (A)11, pp. 143-156.
- LEHMANN, E.L.(1959). Testing statistical hypothesis. J. Wiley and Sons, New-York. 369 p.
- MAC MILLAN, R.G.(1968). Tests for one or two outliers. Ph. D. Thesis., North Carolina State Univ. 100 p. Publi  sur demande par : University Microfilms Limited, High Wycomb, Bucks. Angleterre.
- MAC MILLAN, R.G.(1971). Tests for one or two outliers in normal sample with unknown variance. *Technometrics*, Vol. 13, n° 1 , pp. 87-100.
- MAC MILLAN, R.G.  
MORAN, M.A.(1973). Tests for one or two outliers in normal sample with unknown variance : a correction. *Technometrics*, Vol. 15, n° 3, pp. 637.640.
- METIVIER, M.(1968). Notions fondamentales de la th orie des probabilit s. Dunod, Paris. 304 p.
- MURPHY, R.B.(1951). On tests for outlying observations. Ph. D. Thesis, Princeton University. 101 p. Publi  sur demande par : University microfilms Limited, High Wycomb, Bucks. Angleterre.
- PAULSON, E.(1952). An optimum solution to the k-sample slippage problem for the normal distribution. *A.M.S.* 23, pp. 610-6.
- PEARSON, E.S.  
CHANDRA SEKAR, C.(1936). The efficiency of statistical tools, and a criterion for the rejection of outlying observations. *Biometrika* 28, pp. 308-20.

- PEIRCE, B. (1852).  
Criterion for the rejection of doubtful observations. *Astronomical journal* 2, pp. 161-3.
- QUESENBERRY, C.P.  
DAVID, H.A. (1961)  
Some tests for outliers. *Biometrika* 48, pp. 379-90.
- SARHAN, A.E.  
GREENBERG, B.G. (1962).  
Contributions to order statistics. J. Wiley and Sons, New-York. 482 p.
- SIOTANI, M. (1959).  
The extreme value of the generalised distances of the individual points in the multivariate normal sample. *Ann. Inst. Stat. Math.*, Tokyo 10, pp. 183-208.
- TEICHROE, D. (1956).  
Tables of expected values of order statistics and products of order statistics for sample of size twenty and less from the normal distribution. *A.M.S.* 27. pp. 410-26.
- THOMPSON, W.R. (1935).  
On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *A.M.S.* 6, pp. 214-219.
- ULMO, J. BERNIER, J. (1973).  
Eléments de décision statistique. P.U.F. Paris. 330 p.
- WALD, A. (1950). A  
Statistical decision functions. Chelsea Publ. Company, New-York. 179 p.
- WRIGHT, T.W. (1884).  
A treatise on the adjustment of observations by the method of least squares. New-York. §§ 69-73.
- WILKS, S.S. (1962).  
Mathematical Statistics. J. Wiley and Sons. New-York. 644 p.
- WILKS, S.S. (1963).  
Multivariate statistical outliers. *Sankhya* A 25, pp. 407-26.