N° d'ordre 2764

THESE

présentée

A L'UNIVERSITE PAUL SABATIER DE TOULOUSE

pour obtenir

LE GRADE DE DOCTEUR DE 3ème CYCLE

Spécialité: MATHEMATIQUES APPLIQUEES

par

Joseph SAINT PIERRE

TEST D'HYPOTHESES D'INEGALITE DANS LE MODELE LINEAIRE GENERALISE

Soutenue le [illegible] 1983 devant la Commission d'Examen

MM. H. CAUSSINUS — Président

J.M. BREMNER
O. BUNKE
J.R. MATHIEU
X. MILHAUD
A. RAUGI — Examinateurs

SAINT-PIERRE (Joseph). - Tests d'hypothèses d'inégalité dans le modèle linéaire généralisé. - 60 p.

Th. 3ème cycle : Mathématiques Appliquées : Toulouse III : 1983 ; 2764.

RESUME :

Dans un premier chapitre, on donne quelques rappels indispensables sur les problèmes de statistique inférentielle sous contraintes d'inégalités et notamment sur la régression isotone et des tests unilatéraux dans le cadre de lois normales multidimensionnelles. Le deuxième chapitre est consacré à la présentation du modèle linéaire généralisé multidimensionnel avec un rappel des principaux résultats concernant la normalité asymptotique des estimateurs du maximum de vraisemblance. Dans le troisième chapitre on construit des tests d'une hypothèse simple contre une hypothèse d'inégalité portant sur toutes les composantes du paramètre, ainsi que des tests d'hypothèses d'inégalité portant sur toutes les composantes du paramètre, cela dans le cadre du modèle linéaire généralisé. On montre que les statistiques de ces tests sont asymptotiquement distribuées suivant la loi de $\bar{\chi}^2$. Dans le quatrième chapitre on construit des tests d'une hypothèse linéaire composée contre une hypothèse d'inégalité. On détermine aussi la distribution asymptotique des statistiques de test.

MOTS CLES :

- Régression isotone
- Hypothèses restreintes
- Test du rapport de vraisemblance
- Modèle linéaire généralisé
- Structure gaussienne tangente
- cône convexe polyédrique
- Test des multiplicateurs de Lagrange
- Loi de $\bar{\chi}^2$
- Probabilité orthante

JURY et date de soutenance : 3 Février 1983

Président : H. CAUSSINUS
Membres : J.M. BREMNER
O. BUNKE
J.R. MATHIEU (Laboratoire de Statistique et Probabilités)
X. MILHAUD
A. RAUGI

Je remercie

Henri CAUSSINUS pour avoir accepté de présider le jury de cette thèse, ainsi que pour les nombreux conseils qu'il m'a prodigués lors de l'élaboration de celle-ci,

Jean-René MATHIEU qui a dirigé ce travail, je le remercie pour sa disponibilité et la qualité des relations que j'ai eues avec lui,

Olaf BUNKE, Professeur à l'Université Humboldt de Berlin, pour s'être intéressé à mon travail, pour m'avoir aidé à l'amélioration de celui-ci et pour l'immense honneur qu'il me fait en acceptant de participer au jury de cette thèse,

J.M. BREMNER, Professeur à l'Université de Kent à Canterbury, pour l'intérêt qu'il a manifesté pour mon travail lors de sa venue à COMPSTAT 82, pour avoir lu une version écrite en anglais de cette thèse, et pour l'honneur qu'il me fait en acceptant de se déplacer pour participer au jury de celle-ci,

Xavier MILHAUD pour avoir suivi mes exposés, pour les remarques qu'il m'a faites et sa participation au jury,

Albert RAUGI pour m'avoir aidé dans mon travail, pour avoir accepté de le lire et de participer au jury,

Akio KUDÔ, Professeur à l'Université de Kyushu à Fukuoka pour l'intérêt qu'il a porté à mon travail lors de notre rencontre à Toulouse au cours de COMPSTAT 82 et pour m'avoir fait parvenir ses publications.

Je remercie P. BAYLAC et R. DEBEURRE pour l'efficacité avec laquelle elles ont mené à bien l'ingrat travail de frappe, O. LANDREVIE dont la gentillesse et la compétence permettent de résoudre la plupart des problèmes administratifs, Ch. TERTRE pour le travail d'impression.

Je remercie tous mes copains du laboratoire qui ont contribué à rendre l'ambiance de travail plus agréable, je ne peux les citer tous mais je ne peux m'empêcher deremercier personnellement : Mostafa, Amir, Bébert, Bernard, Moha, Bachir, Lucile, Véro, J.R., Pierre Bru ... j'en ai sans doute oublié ne m'en veuillez pas ...

INTRODUCTION

Considérons la situation suivante : soit deux champs en tout point identiques, on fait dans l'un et l'autre champ la même culture, dans le premier champ on met un engrais fertilisant et dans le second on ne met rien. Il est alors possible d'affirmer a priori que le rendement dans le second champ sera inférieur à celui du premier. Cet exemple de situation conduit le statisticien à l'utilisation du test dit "test unilatéral" s'il veut juger de l'efficacité de l'engrais. La justification de cela est évidente : on jugera en effet que l'engrais est "bon" seulement si le rendement observé dans le premier champ est supérieur à celui du second. Alors que si on veut juger de "l'effet" de l'engrais avec le "test bilatéral" on prend en compte la valeur absolue de la différence des rendements observés. Il convient donc dans chaque situation statistique d'utiliser les informations obtenues "a priori". Ce genre de situations a conduit à la théorie de la régression isotone et aux tests d'hypothèses ordonnées ainsi qu'aux tests "unilatéraux multidimensionnels". Ces théories ont été mises au point dans le modèle linéaire par D.J. Bartholomew (1959), A. Kudô (1963), P. Nüesch (1966) et sont largement commentées dans Barlow et Coll. (1972).

La normalité n'étant pas toujours requise, certains auteurs se sont intéressés à l'extension de ces tests à des situations où les variables observées suivent des lois autres que la loi normale, mais cependant usuelles : binomiale, poisson etc... Dans Barlow et Coll. (1972) certains cas particuliers sont étudiés. T. Robertson et E.J. Wegman (1978) se sont penchés sur le problème des tests d'hypothèses ordonnées dans une structure exponentielle. Les tests de ces auteurs utilisent le critère du rapport de vraisemblance. D'autre part, afin de recouvrir ce grand nombre de loi de probabilités usuelles J. Nelder et R.W.M. Wedderburn (1972) ont introduit le modèle linéaire généralisé qui constitue une extension des structures exponentielles. J.R. Mathieu (1978, 1981) a donné une version du modèle linéaire généralisé multidimensionnelle afin de pouvoir étudier certaines situations n'entrant pas dans le cadre du modèle linéaire généralisé de J.A. Nelder et R.W.M. Wedderburn, notamment les tables de contingences. Les auteurs cités ont étudié dans le modèle linéaire généralisé les tests d'hypothèse linéaires ; le but de notre travail est d'étudier dans ce modèle les tests "unilatéraux multidimensionnels".

Dans le premier chapitre on donne des rappels sur les tests d'hypothèses restreintes dans le modèle linéaire c'est-à-dire sur la régression isotone.

Dans le deuxième chapitre on présente le modèle linéaire généralisé (G.L.I.M.) dans sa version multidimensionnelle, on rappelle les principaux résultats notamment sur la normalité asymptotique des estimateurs.

Dans le troisième chapitre on construit dans le G.L.I.M. des tests d'une hypothèse simple contre une hypothèse imposant au paramètre des contraintes d'inégalités sur toutes ses composantes. On montre que les statistiques de tests sont asymptotiquement distribuées suivant une loi de $\bar{\chi}^2$ que l'on explicite. On construit aussi des tests d'une hypothèse imposant au paramètre des contraintes d'inégalité sur toutes ses composantes contre l'hypothèse contraire, on établit la distribution asymptotique des statistiques de tests. Dans l'un et l'autre cas on utilise la statistique du rapport de vraisemblance ainsi qu'une statistique dont l'écriture est similaire à la statistique du test de Wald d'une hypothèse linéaire. Dans le cas d'une hypothèse nulle imposant au paramètre des contraintes d'inégalités, on construit deux tests à l'aide de multiplicateurs de Lagrange. On montre l'équivalence asymptotique des différents tests utilisés.

Dans le quatrième chapitre on se situe toujours dans le G.L.I.M ; on s'intéresse aux tests d'une hypothèse nulle linéaire composée contre une hypothèse alternative imposant des contraintes d'inégalités au paramètre. La difficulté supplémentaire par rapport au chapitre précédent est due au fait que les distributions asymptotiques des statistiques de tests dépendent de la vraie valeur du paramètre sous l'hypothèse nulle ; comme celle-ci est composée, il est délicat de construire des tests de niveau asymptotiques constant, ce qu'on s'est efforcé de faire ici.

CHAPITRE I

RAPPELS SUR LES TESTS D'HYPOTHÈSES RESTREINTES DANS LE CAS GAUSSIEN

- 0 - 0 - 0 -

I.1 - <u>Tests unilatéraux sans une structure gaussienne</u>

Soit la structure statistique :

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu,\sigma^2) \;,\; \mu \in \mathbb{R} \;,\; \sigma \in \mathbb{R}_+^*\})^n .$$

On s'intéresse au test

$$\begin{aligned} \text{de } H_0 &: (N(\mu_0,\sigma^2) \;,\; \sigma \in \mathbb{R}_+^*) \\ \text{contre } H_1 &: (N(\mu,\sigma^2) \;,\; \mu > \mu_0 \;,\; \sigma \in R_+^*) \;. \end{aligned}$$

POur abrégér les notations on écrit $H_0 : \mu = \mu_0$ et $H_1 : \mu > \mu_0$.

$X_1, X_2,\ldots,X_n$ les applications coordonnées ce sont des variables aléatoires réelles (v.a.r.) indépendantes de loi $N(\mu,\sigma^2)$.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad\qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$\frac{\sqrt{n}(\bar{X}-\mu)}{S}$ suit une loi de Student à (n-1) degrés de liberté.

On notera $t_{\nu,\alpha}$ la valeur qui est dépassée avec une probabilité α par une variable aléatoire de Student à ν degrés de liberté. Le test de Student unilatéral de H_0 contre H_1 est le suivant (test de niveau α).

rejet de H_0 pour $\frac{\sqrt{n}(\bar{X}-\mu)}{S.} > t_{n-1,\alpha}$

c'est sans doute l'exemple le plus simple et le plus connu de statistique inférentielle sous contrainte d'ordre.

Soit la structure statistique

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_1,\sigma^2)\}, \mu_1 \in \mathbb{R} \;,\; \sigma \in \mathbb{R}_+^*\})^{n_1} \otimes (\mathbb{R},\mathcal{B}_{\mathbb{R}}\{N(\mu_2,\sigma^2), \mu_2 \;\mathbb{R}, \sigma \in \mathbb{R}_+^*\})^{n_2}$$

on s'intéresse au test de $H_0 : \mu_1 = \mu_2$

contre $H_1 : \mu_1 > \mu_2$

$\bar{X}_1$ (resp.$\bar{X}_2$) , S_1^2 (resp.S_2^2) désignent les mêmes choses que précedemment dans la structure $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_1,\sigma^2), \mu_1 \in \mathbb{R}, \sigma \in \mathbb{R}_+^*)^{n_1}$ (resp.$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_2,\sigma^2), \mu_2 \in \mathbb{R}, \sigma \in \mathbb{R}_+^*\})^{n_2}$).

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \;,$$

Le test de Student au niveau de H_o contre H_1 se fait de la manière suivante :

$$\text{rejet de } H_o \text{ pour } \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S^2(\frac{1}{n_1} + \frac{1}{n_2})}} > t_{\nu,\alpha} \quad \nu = n_1 + n_2 - 2 \ .$$

C'est un autre exemple élémentaire de statistique inférentielle sous contraintes d'ordre.

Problème :

Soit la structure statistique

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_1,\sigma^2),\ \mu_1 \in \mathbb{R}\})^{n_1} \otimes (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_2,\sigma^2),\ \mu_2 \in \mathbb{R}\})^{n_2}$$
$$\otimes (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_3,\sigma^2), \mu_3\ \mathbb{R}\})^{n_3} \ .$$

Soient les hypothèses

$$H_o : \mu_1 = \mu_2 = \mu_3$$
$$H_1 : (\mu_1 \leq \mu_2 < \mu_3) \cup (\mu_1 < \mu_2 \leq \mu_3) \ .$$

Tester H_o contre H_1 semble être une généralisation naturelle de la situation précédente.
Pourtant cette généralisation entraîne certaines difficcultés. On est obligé d'étendre la théorie des tests unilatéraux au cas multidimensionnel.
Une situation un peu plus générale est la suivante ; on dispose de la structure stztistique :

$$\bigotimes_{i=1}^{k} (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_i,\sigma^2),\ \mu_i \in \mathbb{R}\})^{n_i} \ .$$

Soient les hypothèses

$$H_o : \mu_1 = \mu_2 = \ldots = \mu_k$$
$$H' : \mu_1 \leq \mu_2 \leq \ldots \leq \mu_R$$
$$H_1 : H' \setminus H_o \ .$$

On veut tester H_o contre H_1.

Ces exemples de situation ont conduit à la théorie de la régression isotone. Historiquement ce sont les tests d'hypothèses ordonnées qui ont été étudiés les premiers par Bartholonew. Un exposé général des m^{e}thodes peut être trouvé dans Barlow, Bartholomew, Bremner, Brunk (1972) chapitre III.

Lorsqu'on veut dans l'exemple précédent tester H_0 contre H_1 et que l'on veut utiliser le test du rapport de vraisemblance on a besoin pour le calcul de la statistique de test de l'estimation du maximum de vraisemblance des μ_i sous la contrainte H_0 ce qui est facile, mais on a besoin aussi de l'estimation des μ_i sous la contrainte H_1 ce qui est moins facile. L'estimation des μ_i sous la contrainte H_1 conduit à la régression isotone.

I.2 - La régression isotone (d'après Barlow et Coll (1972)).

Définition : *Soit X un ensemble fini à k éléments.*
$X = \{x_1, x_2, \ldots, x_k\}$ X est muni de l'ordre simple

$$x_1 < x_2 < \ldots < x_k .$$

Une fonction f sur X à valeurs réelles est dite isotone si $x_i, x_j \in X$ et $x_i < x_j$ implique $f(x_i) \leq f(x_j)$.
On pourrait aussi utiliser la terminologie fonction monotone sur X mais les auteurs précités préfèrent le mot isotone.
On note $I(X)$ l'ensemble des fonctions isotones sur X.

Définition : *Soit g une fonction donnée sur X à valeurs réelles et w une fonction donnée sur X à valeurs réelles et positives une fonction g^* isotone sur X est une régression isotone de g avec les poids w par rapport à l'ordre simple $x_1 < x_2 < \ldots < x_k$ si elle vérifie l'égalité suivante*

$$\sum_{x \in X} [g(x) - g^*(x)]^2 \, w(x) = \inf_{f \in I(X)} \sum_{x \in X} [g(x) - f(x)]^2 \, w(x) .$$

c'est-à-dire que g^* est la meilleure approximation isotone de g.

Remarques :

Rechercher g^* est strictement la même chose que rechercher le vecteur de $\mathbb{R}^k$ de composantes $g^*(x_1), g^*(x_2), \ldots, g^*(x_k)$ tel que $g^*(x_1) \leq g^*(x_2) \leq \ldots \leq g^*(x_k)$ et qui minimise la quantité $\sum_{i=1}^{k} [(g(x_i) - y_i]^2 \, w(x_i)$ parmi les vecteurs $y \in \mathbb{R}^k$ tels que $y_1 \leq y_2 \leq \ldots \leq y_k$.

Soit $C = \{ y \in \mathbb{R}^k \;/\; y_1 \leq y_2 \leq \ldots \leq y_k \}$.

Munissons $\mathbb{R}^k$ du produit scalaire $<,>_w$ défini par

$$<y, y'>_w = \sum_{i=1}^{k} y_i \, y'_i \, w(x_i).$$

on note $\| \;\|_w$ la norme associée.

. Le problème de recherche de la regression isotone se ramène au problème de minimisation suivant :

$$\operatorname*{Inf}_{y \in C} \| {}^t[g(x_1),\ldots,g(x_k)] - y \|_w^2 \qquad (1)$$

. C est un cône convexe et fermé (évident).

Proposition I.1 : *La régression isotone d'une fonction g existe et est unique.*

En effet C étant convexe et fermé il résulte directement du théorème de projection que le problème admet une solution unique ${}^t(y\,(x_1),\ldots,g\,(x_k))$ qui est la projection de ${}^t(g(x_1),\ldots,g(x_k))$ sur C au sens du rpoduit scalaire $\;,\;_w$.

I.2.1 - Interprétation graphique

Soit g et w 2 fonctions définies sur X de la même manière que précédemment et considérons un repère orthonormé du plan. Plaçons dans ce repère k+1 points définis par leurs coordonnées P_j $j=0,1,\ldots,k$.

$$P_j = (W_j, G_j)$$

avec $$W_j = \sum_{i=1}^{j} w(x_i)$$

et $$G_j = \sum_{i=1}^{j} w(x_i)\,g(x_i) \qquad \text{pour } j \neq 0$$

$$P_0 = (W_0, G_0) = (0,0)$$

Le diagramme joignant successivement les points P_j depuis P_0 jusqu'à P_k est appelé "diagramme des sommes cumulées" 'en anglais : cumulative Sum Diagram d'où l'abréviation C.S.D.). On considère le "plus grand minorant convexe" (en anglais "greatest convex minorant G.C.M.) du diagramme des sommes cumulées et on note P_j^* le point du G.C.M. d'abscisse W_j et on définit G_j^* comme étant l'ordonnée du point P_j^* $P_j = (W_j, G_j)$.

Proposition I.2 : *En chaque point x_j de X la pente du G.C.M. c'est-à-dire $\dfrac{G_j^* - G_{j-1}^*}{W_j - W_{j-1}}$ est la valeur de la régression isotone de g en x_j, $g^*(x_j)$.*

La démonstration de ce résultat se trouve dans Barlow et Coll. (1972) (théorème I.1 page 112).

L'idée est la suivante :

Si C est un convexe fermé si x^* est la projection de X sur C au sens de $<,>$, alors x^* est caractérisé par la relation

$$\forall\, y \in C \qquad < x - x^* \,,\, x^* - y > \geq 0$$

cette inégalité appliquée à la régression isotone donne :

g^* régression isotone de g si et seulement si $\forall\, f \in I(X)$

$$\sum_{x \in X} [g(x) - f(x)]^2\, w(x) \geq \sum_{x \in X} [g(x) - g^*(x)]^2\, w(x) + \sum_{x \in X} [g^*(x) - f(x)]^2\, w(x).$$

on vérifie que g^* définie par $g^*(x_j) = \dfrac{G_j^* - G_{j-1}^*}{W_j - W_{j-1}}$ satisfait cette inégalité , et d'après l'unicité de la régression isotone on prouve que c'est effectivement la régression isotone de g .

I.2.2. Propriétés de la régression isotone

$\forall$ g fonction de X dans $\mathbb{R}$

$\forall\, \mu \in \mathbb{R} \quad \forall\, \lambda \in \mathbb{R}_+^*$

$(g+\mu)^* = g^* + \mu$

$(\lambda g)^* = \lambda g^*$

La démonstration de ces propriétés est immédiate.

I.2.3. Calcul de la régression isotone.

Il existe plusieurs algorithmes de calcul de la régression isotone assez longuement détaillés dans Barlow et Coll. (1972). Il existe une formule simple pour l'écriture permettant de calculer $g^*(x_i)$ dite formule max-min

$$\forall\, i = 1,\dots,k \qquad g^*(x_i) = \max_{s \leq i} \min_{t \geq i} \frac{\sum_{r=s}^{t} g(x_r) w(x_r)}{\sum_{r=s}^{t} w(x_r)}$$

Remarques :

Ce n'est pas la formule la plus simple ni la plus rapide pour calculer $g^*(x_i)$.

Il existe des formules équivalentes par exemple on a

$$g^*(x_i) = \min_{t \geq i} \max_{s \leq i} \frac{\sum_{r=s}^{t} g(x_r)\, w(x_r)}{\sum_{r=s}^{t} w(x_r)}$$

I.2.4. Estimation du maximum de vraisemblance sous contraintes d'ordre dans une structure gaussienne.

Soit la structure statistique suivante

$$\bigotimes_{i=1}^{k} (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_i,\sigma_i^2),\ \mu_i \in \mathbb{R}\})^{n_i}$$

on note X_{ij} les applications coordonnées. X_{ij} sont des variables aléatoires réelles indépendantes de loi $N(\mu_i,\sigma_i^2)$.

On suppose que les σ_i^2 sont connus

$$\sum_{i=1}^{k} n_i = n$$

La vraisemblance s'écrit

$$\frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{k} \frac{1}{\sigma_i^{n_i}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \right\}$$

maximiser la vraisemblance se ramène à minimiser

$$\sum_{i=1}^{k} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \qquad (2)$$

en posant $\bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$

on trouve

$$(2) = \sum_{i=1}^{k} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 + \sum_{i=1}^{k} \frac{n_i}{\sigma_i^2} (\bar{x}_{i.} - \mu_i)^2$$

maximiser la vraisemblance se ramène donc à minimiser

$$\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2} (\bar{x}_{i.} - \mu_i)^2$$

On pose $w_i = \frac{n_i}{\sigma_i^2}$.

Soit H_0 la contrainte $\mu_1 = \mu_2 = \dots = \mu_k$.

Il est bien connu que la valeur de μ qui minimise

$$\sum_{i=1}^{k} w_i(\bar{x}_{i.} - \mu)^2 \quad \text{est} \quad \hat{\mu} = \frac{\sum_{i=1}^{k} w_i \bar{x}_{i.}}{\sum_{i=1}^{k} w_i} .$$

Sous la contrainte H_o l'estimateur du maximum de vraisemblance de μ est donc

$$\bar{X} = \frac{\sum_{i=1}^{k} w_i \bar{X}_{i.}}{\sum_{i=1}^{k} w_i} \quad \text{où} \quad \bar{X}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

Soit H' la contrainte $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$.

On veut minimiser $\sum_{i=1}^{k} w_i(\bar{x}_{i.} - \mu_i)^2$ sous H' .

D'après ce que l'on a vu les μ_i^* qui vérifient

$$\sum_{i=1}^{k} w_i(\bar{x}_{i.} - \mu_i^*)^2 = \operatorname*{Inf}_{\mu \in H'} \sum_{i=1}^{k} w_i(\bar{x}_{i.} - \mu_i)^2$$

constituent ce que l'on appelle régression isotone des $\bar{x}_{i.}$ avec les poids w_i.

I.3. <u>Test du rapport de vraisemblance dans le cas normal d'une hypothèse d'homogénéité contre une hypothèse ordonnée des moyennes.</u>

1.3.1. <u>Test du $\bar{\chi}^2$</u>

Soit la structure statistique $\bigotimes_{i=1}^{k} (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_i, \sigma_i^2), \mu_i \in \mathbb{R}\})^{n_i}$

$H_o \quad \mu_1 = \mu_2 = \ldots = \mu_k$

$H' \quad \mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$

$H_1 = H' \setminus H_o$

Soit λ le rapport de vraisemblance

$$\lambda = \frac{\max_{\mu \in H'} \text{ vraisemblance}}{\max_{\mu \in H_o} \text{ vraisemblance}}$$

$$T = 2 \operatorname{Log} \lambda = \sum_{i=1}^{k} w_i(\bar{x}_{i.} - \hat{\mu})^2 - \sum_{i=1}^{k} w_i(x_{i.} - \mu_i^*)^2 \quad .$$

Soit C le cône de $\mathbb{R}^k$ déjà défini $C = \{x \in \mathbb{R}^k / x_1 \leq x_2 \leq \ldots \leq x_k\}$ comme on l'a vu ${}^t(\mu_1^*, \ldots, \mu_k^*)$ est la projection du vecteur ${}^t(\bar{x}_{1.}, \ldots, \bar{x}_{k.})$ sur C au sens du produit scalaire déjà défini.

C est un cône polyèdrique. Le vecteur $\hat{\underline{\mu}}$ (${}^t\hat{\underline{\mu}} = (\hat{\mu},\ldots,\hat{\mu})$) appartient à toutes les faces de ce polèdre. On applique le théorème de pythagore à la face de ce polyèdre (plus précisément au sous-espace vectoriel qui porte cette face) sur laquelle se trouve μ^* (${}^t\mu^* = (\mu_1^*,\ldots \mu_k^*)$). On trouve

$$\sum_{i=1}^{k} w_i(\bar{x}_{i.} - \hat{\mu})^2 - \sum_{i=1}^{k} w_i(\bar{x}_{i.} - \hat{\mu}_i^*)^2 = \sum_{i=1}^{k} w_i(\mu_i^* - \hat{\mu})^2$$

La statistique du test du rapport de vraisemblance de H_0 contre H_1 est donc $\sum_{i=1}^{k} w_i(\mu_i - \hat{\mu})^2$.

A cause de la similitude d'écriture avec le test du rapport de vraisemblance de H_0 contre $\mathbb{R}^k \setminus H_0$ les premiers auteurs à avoir utilisé cette statistique l'ont appelée statistique de $\bar{\chi}_k^2$; remarquons tout de suite que le paramètre k ne caractérise pas la distribution de la statistique sous l'hypothèse nulle car celle-ci dépend aussi de H et des poids $w_1, w_2,\ldots,w_k$.

Le test du rapport de vraisemblance de H_0 contre H_1 se fait de la manière suivante : rejet de H_0 pour les grandes valeurs de $\bar{\chi}_k^2$.

<u>Remarque</u> : dans le cas usuel où les variances sont égales on a l'expression

$$\sigma^2 \bar{\chi}_k^2 = \sum_{i=1}^{k} n_i\ (\mu_i^* - \hat{\mu})^2$$

I.3.2. <u>Test du $\bar{E}^2$</u>

Dans la structure statistique $\bigotimes_{i=1}^{k} (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu_i,\sigma_i^2),\ \mu_i \in \mathbb{R},\ \sigma_i \in \mathbb{R}_+\})^{n_i}$

si les σ_i sont inconnus mais peuvent êtremis sous la forme $\sigma_i = a_i\ \sigma^2$ où les a_i sont connus et σ^2 inconnu on peut étendre dans ce cas là le test du $\bar{\chi}^2$. En particulier dans le cas où les $a_i = 1 \quad \forall\ i=1,\ldots,k$.

Pour tester H_0 contre H_1 la statistique du test du rapport de vraisemblance est

$$\bar{E}_k^2 = \sum_{i=1}^{k} n_i(\mu_i - \hat{\mu})^2 \Big/ \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu})^2$$

on rejette H_0 pour les grandes valeurs de $\bar{E}_k^2$.

Pour construire les tests du $\bar{\chi}^2$ ou du $\bar{E}^2$ il est intéressant d'avoir la distribution des statistiques de test sous l'hypothèse nulle afin de déterminer le niveau des tests construits ou pour savoir comment construire un test de niveau donné. Nous allons donc énoncer 2 résultats qui déterminent les distributions recherchées. Ces résultats se trouvent avec leurs démonstration dans Barlow et Coll. (1972).

Proposition I.3 : *Si H_0 est vraie*

$$\forall c > 0 \qquad P(\bar{\chi}_k^2 \geq c) = \sum_{\ell=2}^{k} p(\ell,k)\ P(\chi_{\ell-1}^2 \geq c)$$

$$P(\bar{\chi}_k^2 = 0) = p(1,k)$$

où $p(\ell,k)$ désigne la probabilité que la régression isotone M_i^* prenne exactement ℓ valeurs distinctes et χ_ν^2 désigne une variable aléatoire de χ^2 à ν dégrés de liberté ce qui justifie aussi a posteriori la notation $\bar{\chi}^2$.

Proposition I.3 bis :

Si H_0 est vraie

$$\forall c > 0 \qquad P(\bar{E}_k^2 \geq c) = \sum_{\ell=2}^{k} p(\ell,k)\ P(B_{\frac{1}{2}(\ell-1),\frac{1}{2}(n-\ell)} \geq c)$$

$$P(\bar{E}_k^2 = 0) = p(1,k)$$

où $p(\ell,k)$ a la même signification que dans la proposition précédente et $B_{a,b}$ désigne une variable aléatoire ayant la distribution bêta de paramètres a et b.

Remarque :

Les distributions du $\bar{\chi}_k^2$ et du $\bar{E}_k^2$ dépendent des coefficients $p(\ell,k)$, ces coefficients dépendent de l'ordre et des poids utilisés pour la régression isotone.

La difficulté de la mise en oeuvre des tests résulte de la difficulté du calcul de ces coefficients.

I.3.3 - Calcul des $p(\ell,k)$ dans le cas de l'ordre simple et des poids égaux.

Dans ce cas particulier on a le résultat suivant

$$p(\ell,k) = \frac{|S_k^\ell|}{k!}$$

où $|S_k^\ell|$ désigne le coefficient de z^ℓ dans le polynôme

$$z(z+1)\ \dots\ (z+k-1) = \prod_{j=0}^{k-1} (z+j).$$

Les $p(\ell,k)$ vérifient

$$p(1,k) = \frac{1}{k}$$

$$p(k,k) = \frac{1}{k!}$$

$$p(\ell,k) = \frac{1}{k}\, p(\ell-1,k-1) + \frac{k-1}{k}\, p(\ell,k-1) \ .$$

Il n'est pas difficile, à partir de ces formules, d'établir des tables de coefficients $p(\ell,k)$ surtout lorsque k n'est pas trop grand.
Pour les tables de coefficients $p(\ell,k)$ et pour le calcul de ces coefficients dans les cas où l'ordre n'est pas simple et où les poids ne sont pas égaux on peut encore se référer à Barlow et coll. (1972).

I.4 - Equivalent du test unilatéral dans le cas multidimensionnel

Le problème est le suivant : $Y_1, Y_2,\dots,Y_n$ sont n variables aléatoires à valeurs dans $\mathbb{R}^k$ indépendantes et de même loi Y_i suit une loi $N_k(\mu,\Sigma)$ $\mu \in \mathbb{R}^k$ et Σ est la matrice des variances et covariances des Y_i.

Les Y_i sont des variables aléatoires multinormales.

On considère les hypothèses suivantes

H_0 $\mu = 0$ i.e. $\mu_i = 0$ $\forall$ i=1,...,k

H $\mu_i \geq 0$ $\forall$ i=1,...,k

$H_1 = H \setminus H_0$.

Et on s'intéresse au test de l'hypothèse H_0 contre l'hypothèse H_1.

Remarque :

Par changement de variables de nombreux problèmes peuvent se ramener à l'étude de ce test, c'est notamment le cas des tests que l'on fait dans le cadre de la régression isotone.

Peut se ramener au problème précédent la situation suivante :
g est une bijection linéaire de $\mathbb{R}^k$ dans $\mathbb{R}^k$

H_0 : $\mu = 0$

H : $g(\mu) \geq 0$ (c'est-à-dire toutes les composantes de $g(\mu)$ sont positives ou nulles)

H_1 : $H \setminus H_0$

et on veut tester H_0 contre H_1.

Ce type de tests ont été étudiés simultanément par Kudô (1963) et Nüesch (1964) en étant considérés par ces auteurs comme une généralisation des tests d'hypothèses ordonnées étudiés par Bartholomew (1959).

On a la structure statistique suivante

$$(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}, \{N_k(\mu,\Sigma), \mu \in \mathbb{R}^k\})^n$$

Σ est supposée connue symétrique définie positive.

On utilise le test du rapport de vraisemblance.

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i .$$

$\bar{Y}$ est l'estimateur du maximum de vraisemblance de μ dans la structure précédente. La statistique du test du rapport de vraisemblance que l'on appelle aussi $\bar{\chi}^2$ est donnée par

$$\bar{\chi}^2 = n\{ {}^t\bar{Y}\,\Sigma^{-1}\,\bar{Y} - \min_{\mu_i \geq 0} {}^t(\bar{Y}-\mu)\Sigma^{-1}(\bar{Y}-\mu)\}.$$

On est donc conduit au problème de minimisation suivant

$$\min_{\mu_i \geq 0} {}^t(\bar{Y}-\mu)\,\Sigma^{-1}(\bar{Y}-\mu)$$

Soit H le cône convexe polyèdrique défini par

$$H = \{x \in \mathbb{R}^k \,/\, x_i \geq 0 \quad \forall\, i=1,\dots,k\}.$$

On considère le produit scalaire induit par Σ^{-1}

$$\forall\, x,y \quad \mathbb{R}^k \qquad \langle x,y \rangle_{\Sigma^{-1}} = {}^t x\,\Sigma^{-1}\,y .$$

On note $\|.\|_{\Sigma^{-1}}$ la norme associée.

Le problème s'écrit de la manière suivante :

trouver $\mu^* \in H$ tel que

$$\|\bar{Y} - \mu^*\|_{\Sigma^{-1}} = \operatorname*{Inf}_{\mu \in H} \|\bar{Y} - \mu\|^2_{\Sigma^{-1}} .$$

On en déduit donc que μ^* est la projection de $\bar{Y}$ sur le convexe fermé H, il en résulte l'existence et l'unicité de μ^*. C'est l'estimateur du maximum de vraisemblance de μ sous la contrainte H

$$\begin{aligned}\bar{\chi}^2 &= n\{ {}^t\bar{Y}\,\Sigma^{-1}\,\bar{Y} - {}^t(\bar{Y}-\mu^*)\,\Sigma^{-1}(\bar{Y}-\mu^*)\} \\ &= n(\|\bar{Y}\|^2_{\Sigma^{-1}} - \|\bar{Y}-\mu^*\|^2_{\Sigma^{-1}})\end{aligned}$$

H est un polyèdre, μ^* est la projection de $\bar{Y}$ sur H, c'est donc la projection sur une face de ce polyèdre considérons le sous-espace vectoriel qui porte cette face en appliquant le théorème de pythagore , on a

$$\|\bar{Y}\|^2_{\Sigma^{-1}} = \|\bar{Y} - \mu^*\|^2_{\Sigma^{-1}} + \|\mu^*\|^2_{\Sigma^{-1}}$$

d'où

$$\bar{\chi}^2 = n\,\|\mu^*\|^2.$$

Pour la mise en oeuvre du test on voit donc qu'il est nécessaire de calculer μ^* et cela nécessite l'utilisation d'algorithmes efficaces de calcul de la projection d'un point sur un cône convexe polyèdrique. Voir par exemple Kudô (1975) Bremner (1982). Soit H le polyèdre déjà défini plus haut . On peut, si k est petit, considérer toutes les faces de H et les sous espaces de $\mathbb{R}^k$ qui portent ces faces. Il y a en tout 2^k faces. Pour calculer la projection d'un point x sur H, on peut utiliser la méthode suivante : calculer la projection de x sur chacun des sous espaces défini (cela est classique) puis, parmi toutes les protions ainsi obtenues, on considère celles qui appartiennent effectivement à H et parmi celles-là on considère celle qui minimise la distance voulue .
Dès que k est assez grand cette méthode peut s'avérer assez lourde et il faut donc trouver des méthodes plus performantes, trouver des critères rapides pour déterminer sur quelle face de H se trouve la projection.
Comme dans le cas des hypothèses ordonnées on fait le test de H_0 contre H_1 de la manière suivante :

rejet de H_0 pour les grandes valeurs de $\bar{\chi}^2$

Ici aussi il est intéressant de disposer de la distribution de la statistique $\bar{\chi}^2$ sous hypothèse nulle afin de pouvoir construire des tests de niveau α donné.
On énonce donc un résultat qui se trouve dans Barlow et Coll. (1972).

<u>Proposition I.4</u> : *Si H_0 est vraie*

si $c > 0 \quad P(\bar{\chi}^2 \geq c) = \sum_{j=1}^{k} q(j,k)\; P(\chi_j^2 \geq c)$

$$P(\bar{\chi}^2 = 0) = q(0,k)$$

$q(j,k)$ désigne la probabilité que μ^* ait exactement j coordonnées non nulles.
χ_j^2 dénote une variable aléatoire suivant une loi de χ^2 à j degrés de liberté.

Remarques :

- On peut noter la similitude du rôle joué par $q(j,k)$ dans cette proposition avec celui joué par les $p(\ell,k)$ dans les propositions I.3 et I.3 bis ; cela justifie a posteriori la notation commune $\bar{\chi}^2$ pour les 2 statistiques.
- On a $\sum_{j=o}^{k} q(j,k) = 1$.
- On peut parlerde loi de $\bar{\chi}^2$ on appelle ainsi les "mélanges" de loi de χ^2 c'est-à-dire que la fonction de répartition d'une loi de $\bar{\chi}^2$ est une combinaison linéaire convexe de lois de χ^2.
- Une loi de $\bar{\chi}^2$ n'admet pas a priori une densité par rapport à la mesure de Lebesgue car en général $P(\bar{\chi}^2 = 0) = q(0,k) \neq 0$; la mesure de probabilité admet donc une partie atomique.
- La loi de $\bar{\chi}^2$ obtenue dépendant étroitement de la matrice Σ on introduit donc la notation $\bar{\chi}^2_\Sigma$ pour désigner cette loi

Une forme plus utile de ce résultat est la proposition suivante

Proposition I.5 : *Soit Y une variable aléatoire à valeur dans $\mathbb{R}^k$.*
Y suivant une loi multinormale de moyenne nulle de matrice de variance et covariance Σ connue symétrique définie positive

$$Y \sim N_k(0,\Sigma)$$

$$H = \{ x \in \mathbb{R}^k / x_i \geq 0 \quad \forall\, i=1;\ldots,k \}$$

H cône convexe polyèdrique particulier appelé quadrant positif
Y_H désigne la projection de Y sur H au sens du produit scalaire défini par Σ^{-1}.
$\|Y_H\|^2_{\Sigma^{-1}} = {}^tY_H\, \Sigma^{-1}\, Y_H$ norme carrée de ce vecteur pour le même produit scalaire.
Alors $\|Y_H\|^2_{\Sigma^{-1}}$ suit une loi de $\bar{\chi}^2_\Sigma$.

Pour la démonstration de ces résultats on se réferera à Kudô (1963) et à Nüesch (1966).

I.5 - Test de H contre $\mathbb{R}^k \setminus H$.

On considère encore la structure statistique suivante

$$(\mathbb{R}^k , \mathcal{B}_{\mathbb{R}^k} , \{N_k(\mu,\Sigma) \ , \mu \in \mathbb{R}^k \})^n$$

on considère l'hypothèse H comme précédemment

$$H : \mu \geq 0 \quad \text{c'est-à-dire} \quad \mu_i \geq 0 \quad \forall\ i=1,\dots,k$$

et on s'intéresse au test de l'hypothèse H contre l'hypothèse $\mathbb{R}^k \setminus H$.

Par un calcul analogue à celui fait dans le cas du test de l'hypothèse H_0 contre l'hypothèse H on trouve l'expression de la statistique du rapport de vraisemblance.

On appelle aussi cette statistique $\bar{\chi}^2$ et on a

$$\bar{\chi}^2 = n \, \|\bar{Y} - \mu^*\|^2_{\Sigma^{-1}}$$

$\bar{Y}$ désigne comme précédemment la moyenne des observations c'est-à-dire l'estimateur du maximum de vraisemblance de μ sans contrainte, μ^* est la projection de $\bar{Y}$ sur H au sens de Σ^{-1} c'est-à-dire l'estimateur du maximum de vraisemblance de μ sous la contrainte H.

Ce type de test a été étudié dans le cadre de la régression isotone et des tests d'hypothèse ordonnées par Robertson et Wegman (1978).

On construit le test de la manière suivante

rejet de H pour les grandes valeurs de $\bar{\chi}^2$.

Le test ainsi construit n'est évidemment pas de niveau constant cela étant dû à la complexité de l'hypothèse nulle, comme les auteurs précités on utilisera donc la notion de sous hypothèse la moins favorable, c'est-à-dire celle qui donne la plus grande erreur de première espèce.

$$H_0 : \mu = 0 \qquad H_0 \subset H$$

On va montrer que H_0 est la sous hypothèse la plus défavorable de H dans le sens précisé plus haut. Pour cela on a besoin d'un résultat préliminaire que l'on peut retrouver par exemple dans Robertson et Wegman (1978).

<u>Proposition I.6</u> - *$\forall\ x \in \mathbb{R}^k$ $\forall\ z \in H$ x_H désignant la projection de x sur H au sens d'un produit scalaire quelconque $<,>$, $\|.\|$ désignant la norme associée à ce produit scalaire.*
$(x+z)_H$ désigne la projection de $(x+z)$ sur H au sens de $<,>$ alors

$$\|x+z - (x+z)_H\| \leq \|x - x_H\|$$

Démonstration :

H convexe fermé $(x+z)_H$ projection de x+z sur H est donc caractérisé par
$\forall\, y \in H$, $\|x+z-(x+z)_H\| \le \|x+z-y\|$.

H cône convexe si $z \in H$, $x_H \in H$ alors $x_H + z \in H$ en faisant dans l'inégalité précédente $y = x_H + z$ on trouve le résultat cherché.

Sous H_0 $\quad \bar{Y}$ suit une $N(0, \frac{1}{n}\Sigma)$

Sous μ $\quad \bar{Y}$ suit une $N(\mu, \frac{1}{n}\Sigma)$

Sous H_0 $\quad \bar{Y}+\mu$ suit une $N(\mu, \frac{1}{n}\Sigma)$

$\bar{Y}_H = \mu^*$ le produit scalaire utilisé en l'occurence étant celui induit bien sûr par Σ^{-1}.

Soit $\mu \in H$ d'après la proposition I.6 on a

$$\|\bar{Y}+\mu-(\bar{Y}+\mu)_H\|^2_{\Sigma^{-1}} \le \|\bar{Y}-\mu\|^2_{\Sigma^{-1}}$$

De plus sous μ, $\|\bar{Y}-\mu^*\|^2_{\Sigma^{-1}}$ suit la même loi que

$$\|\bar{Y}+\mu-(\bar{Y}+\mu)_H\|^2_{\Sigma^{-1}} \quad \text{sous } H_0$$

on en déduit donc la proposition suivante

Proposition I.7 : $\forall \; \mu \in H$

$$P_\mu(n\,\|\bar{y}-\mu^*\|^2_{\Sigma^{-1}} \ge c) \le P_{H_0}(n\,\|\bar{y}-\mu^*\|^2_{\Sigma^{-1}} \ge c)$$

Ceci prouve que le test est de niveau maximal pour H_0 et justifie le fait que l'on étudie le test sous l'hypothèse H_0.

Robertson et Wegman (1978) ont montré un résultat analogue pour le cas des hypothèses d'isotonie.

On cherche des tests dont le niveau soit connu sous H_0.

Pour cela on a besoin de connaître la distribution de la statistique $\bar{\chi}^2$ sous H_0.

Proposition I.8 - *Si H_0 est vraie*

$$P(\bar{\chi}^2 = 0) = q(k,k)$$

pour $c>0$

$$P(\bar{\chi}^2 \ge c) = \sum_{\ell=0}^{k-1} q(\ell,k)\, P(\chi^2_{k-\ell} \ge c)$$

où $q(\ell,k)$ désigne comme précédemment la probabilité que μ^* ait exactement ℓ éléments non nuls et χ^2_j désigne une variable aléatoire de χ^2 à j degrés de liberté.

Remarque :

La loi de $\bar{\chi}^2$ ainsi obtenue dépend aussi de la matrice Σ mais comme elle est différente de la loi que nous avons notée $\bar{\chi}^2_\Sigma$ on la notera $\bar{\chi}^2_{(\Sigma^o)}$ et l'on appellera loi polaire de la loi $\bar{\chi}^2_\Sigma$ car la loi $\bar{\chi}^2_\Sigma$ est obtenue en prenant la norme de la projection d'une loi normale sur un cône et la loi $\bar{\chi}^2_{(\Sigma^o)}$ est obtenue en projetant la même loi sur le cône polaire du précédent.

I.6 - Expression de la loi de $\bar{\chi}^2_\Sigma$

Comme dans le cas des tests d'hypothèses d'isotonie on avait besoin de connaîtrr l'expression des coefficients $p(\ell,k)$ on a besoin maintenant, pour déterminer complètement la loi $\bar{\chi}^2_\Sigma$, de connaître l'expression des coefficients $q(j,k)$.

Définition : *Soit Λ une matrice carrée $k \times k$ définie positive et symétrique on appelle probabilité orthante pour la loi normale $N(0,\Lambda)$, et on la note $P(\Lambda)$, la probabilité qu'une variable aléatoire multinormale de moyenne nulle et de matrice de variances et covariances Λ ait toutes ses composantes positives.*

$$P(\Lambda) = \int_H \frac{1}{(\det \Lambda)^{1/2} (\sqrt{2\pi})^n} \exp - \frac{1}{2} {}^t x \, \Lambda^{-1} x \, dx = \int_H d(N(0,\Lambda)).$$

H désigne le quadrant positif de $\mathbb{R}^k$.

Soit $K = \{1,\dots,k\}$

Soit $M \subset K$ on note $M' = K \setminus M$

$n(M)$ = Cardinal de M

Σ_M est la matrice des variances et covariances des Y_i $i \in M$

$\Sigma_{M,M'}$ désigne la même chose sous la condition $Y_j = 0$ si $j \in M'$

On a le résultat important suivant :

$$\mathrm{Proba}(\bar{\chi}^2_\Sigma \geq c) = \sum_{\emptyset \subseteq M \subseteq K} \mathrm{Proba}(\chi^2_{n(M)} \geq c) \; P((\Sigma_{M'})^{-1}) \; P(\Sigma_{M,M'})$$

Ce résultat est dû à A. Kudô (1963).

De là on peut déduire de façon immédiate la valeur des coefficients $q(j,k)$.

Pour la mise en oeuvre pratique des tests le calcul de la probabilité orthante soulève quelques difficultés en effet il n'existe pas de méthodes exactes de calcul pour les dimensions de k supérieures à 3.

I.7 - Lois de $\bar{\chi}^2$ décentrées

Pour l'étude complète des tests qui ont été définis il est nécessaire d'étudier leur puissance pour cela on a besoin des lois de $\bar{\chi}^2$ décentrées.
Soit Y une variable aléatoire à valeur dans $\mathbb{R}^k$.
Y suivant une loi multinormale de moyenne h et de matrice de variance et covariance Σ connue symétrique définie positive

$$Y \sim N_k(h, \Sigma)$$

H est le quadrant positif de $\mathbb{R}^k$, alors $\| Y_H \|^2_{\Sigma^{-1}}$ suit la loi que nous noterons $\bar{\chi}^2_{\Sigma,h}$ loi de $\bar{\chi}^2$ décentrée dépendant de la matrice Σ et du paramètre de décentrage h.
Expression de cette loi.

Il n'est pas possible de donner une expression simple de la loi $\bar{\chi}^2_{\Sigma,h}$.

CHAPITRE II

MODÈLE LINÉAIRE GÉNÉRALISÉ

- o - o - o - o -

II.1 - Introduction

Ce modèle a été introduit par Nelder et Wederburn (1972), il contient le cas des distributions binomiales, des lois de Poisson et les lois gamma ainsi bien sûr que les lois normales.

Ce modèle a été abondamment étudié depuis ; notamment par J.R. Mathieu (1978, 1981) dont nous utilisons les notations. Pour une étude plus détaillée nous renvoyons à J.R. Mathieu (1978, 1981).

On dispose d'un "échantillon" $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ tel que

. ω_i est une réalisation d'une variable aléatoire réelle Ω_i

. ces n variables aléatoires sont indépendantes.

. Ω_i admet une densité $f_i(.,\theta,\xi)$ par rapport à une mesure $\mathcal{M}$ sur $\mathbb{R}$, cette densité est non nulle sur une partie U de $\mathbb{R}$ indépendantes de paramètres inconnus θ et ξ.

. $\forall u \in U \quad f_i(u,\theta,\xi) = \exp\{\alpha(\xi)[\gamma_i u - \mu(\gamma_i) - \eta_i(u)] + \psi(u,\xi)\}$

où t_i est une application linéaire connue de $\mathbb{R}^m$ dans $\mathbb{R}^r$ (1)

ϕ est une application connue de $\mathbb{R}^r$ dans $\mathbb{R}$

θ appartient à $\Theta \subset \mathbb{R}^m$

$\alpha(\xi)$ est strictement positif

μ est une application connue de $\mathbb{R}$ dans $\mathbb{R}$

ψ est une application connue de $\mathbb{R} \times \Xi$ dans $\mathbb{R}$

η_i est une application connue de $\mathbb{R}$ dans $\mathbb{R}$

$$\theta \in \Theta \subset \mathbb{R}^m \xrightarrow{t_i} t_i\theta \in \mathbb{R}^r \xrightarrow{\phi} \gamma_i = \phi(t_i\theta) \in \mathbb{R}$$

Sous des conditions de régularité on peut montrer que Ω_i admet pour moyenne et variance

$$\begin{cases} E_{\theta,\xi}(\Omega_i) = \mu'\circ\phi(t_i\theta) = \mu'(\gamma_i) \\ \sigma^2_{\theta,\xi}(\Omega_i) = \dfrac{\mu''\circ\phi(t_i\theta)}{\alpha(\xi)} \end{cases} \qquad (2)$$

$\alpha(\xi)$ est un paramètre d'échelle puisqu'il n'affecte pas la moyenne, il sera considéré comme un paramètre fantôme ; de plus ce paramètre est souvent connu et alors f_i appartient à la famille exponentielle, le cas le plus important dans lequel $\alpha(\xi)$ est inconnu est le cas d'une loi gaussienne avec une variance σ^2 inconnue qui est alors $1/\alpha(\xi)$.

On s'intéresse donc uniquement au paramètre θ qui est le seul paramètre qui intervient pour la moyenne et dont le rôle est comparable à celui des coefficients d'une régression ; en particulier si $\mu' \circ \phi$ est la fonction identité et si μ'' est une fonction constante on retrouve le modèle classique de la régression linéaire.

II.2 - Modèle linéaire généralisé multidimensionnel

Le modèle précédent ne recouvrant que le cas où les observations sont unidimensionnelles, J.R. Mathieu (1978, 1981) a introduit une version du modèle linéaire généralisé qui recouvre des cas où l'on a affaire avec des observations multidimensionnelles. La situation est la suivante :

On dispose d'un "échantillon" $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ où

. ω_i est la réalisation d'une variable aléatoire à valeur dans $\mathbb{R}^k$, Ω_i

. ces n variables aléatoires sont indépendantes.

. Ω_i admet une densité $f_i(.,\theta,\xi)$ par rapport à une mesure $\mathfrak{m}$ sur $\mathbb{R}^k$, cette densité est non nulle sur une partie U de $\mathbb{R}^k$ indépendante des paramètres inconnus θ et ξ

. $f_i(.,\theta,\xi)$ est définie sur U par

$$\forall u \in U \quad f_i(u,\theta,\xi) = \exp\{\alpha(\xi)[\langle u,\phi(t_i\theta)\rangle - \mu(\gamma_i) - \eta_i(u)] + \phi(u,\xi)\} \qquad (3)$$

avec

. ξ est inconnu $\xi \in \Xi$ $\alpha(\xi)$ est strictement positif.

. θ est inconnu $\theta \in \Theta \subset \mathbb{R}^m$ et Θ est ouvert

. t_i est une application linéaire connue de $\mathbb{R}^m$ dans $\mathbb{R}^r$

. ϕ est une application connue de $\mathbb{R}^r$ dans $\mathbb{R}^{k*}$ où $\mathbb{R}^{k*}$ désigne le dual de $\mathbb{R}^k$

. μ est une application connue de $\mathbb{R}^{k*}$ dans $\mathbb{R}$

. ψ est une application connue de $\mathbb{R}^k \times \Xi$ dans $\mathbb{R}$

. η_i est une application connue de $\mathbb{R}^k$ dans $\mathbb{R}$

. $\langle .,.\rangle$ désigne la dualité entre $\mathbb{R}^k$ et $\mathbb{R}^{k*}$ c'est-à-dire que $\langle u,\phi(t_i\theta)\rangle$ est l'image de l'élément u de U par la forme linéaire $\phi(t_i\theta)$

ainsi nous avons

$$\theta \in \Theta \subset \mathbb{R}^m \xrightarrow{t_i} \beta_i : t_i\theta \in \mathbb{R}^r \xrightarrow{\phi} \gamma_i = \phi(t_i\theta) \in \mathbb{R}^{k*} \xrightarrow{\mu} \mu(\gamma_i) \in \mathbb{R}.$$

On dispose de la structure statistique

$$\bigotimes_{i=1}^{n} \{\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}, f_i(.,\theta,\xi), \theta \in \Theta, \xi \in \Xi\} \qquad (4)$$

Cette structure a été étudiée par J.R. Mathieu (1978, 1981). On peut y trouver des exemples.

Le but de notre travail est d'étudier, dans cette structure statistique, des tests d'hypothèses restreintes sur le paramètre θ semblables aux tests que l'on a présentés dans le chapitre I dans le cadre de la normalité.

Dans le chapitre I on a vu que les tests se faisaient sur le paramètre représentant la moyenne. On a vu aussi dans le cadre du modèle linéaire généralisé unidimensionnel que la moyenne ne dépendait que du paramètre θ il en est de même pour le G.L.I.M. multidimensionnel, cela justifie le fait que l'on s'intéresse aux tests sur θ.

Dans la structure (4) on calcule le logarithme de la vraisemblance et on obtient

$$L_n(\omega,\theta,\xi) = \alpha(\xi) \sum_{i=1}^{n} [\langle\omega_i,\phi(t_i\theta)\rangle - \mu(\gamma_i)-\eta_i(\omega_i)]+ \sum_{i=1}^{n} \psi(\omega_i,\xi). \qquad (5)$$

Afin de mener à bien l'étude voulue nous donnons des rappels de résultats concernant le G.L.I.M.

II.3 - Notations - Hypothèses

II.3.1 - Afin d'étudier la "puissance" asymptotique des tests, on considère la suite $\{\theta_n\}_{n\in N}$ de valeurs du paramètre θ définie par :

$$\theta_n = \overline{\theta} + \frac{h_n}{\sqrt{n}} \qquad (6)$$

où $h_n \to h$

On note P_n et $P_{n'}$ les mesures de probabilités de densités respectives

$$\prod_{i=1}^{n} f_i(.,\overline{\theta},\xi) \quad \text{et} \quad \prod_{i=1}^{n} f_i(.,\theta_n,\xi).$$

Pour établir les propriétés asymptotiques des estimateurs du maximum de vraisemblance on utilise un développement de Taylor de $L_n(.,\theta,\xi)$ dans un voisinage de $\overline{\theta}$, on suppose donc l'existence des dérivées de $L_n(.,\theta,\xi)$ par rapport à θ jusqu'à l'ordre 3. On les note $L_n'(.,\theta,\xi)$, $L_n''(.,\theta,\xi)$. $L_n'''(.,\theta,\xi)$. De plus on leur impose d'être,dans un voisinage de θ , suffisamment petites en norme.

D'autre part on impose à densité f_i des conditions telles que l'on puisse déduire du théorème de Liapounov la convergence en loi de $\frac{L_n'(.,\overline{\theta},\xi)}{\sqrt{n}}$ vers une variable aléatoire gaussienne dont la matrice des variances et covariances sera la limite en probabilité de $-\dfrac{L''_n(.,\theta,\xi)}{n}$; on impose donc à cette limite d'être une forme quadratique positive et non dégénérée.

II.3.2 - Expressions de $L'_n(.,\theta,\xi)$, $L''_n(.,\theta,\xi)$, $L'''_n(.,\theta,\xi)$

Pour simplifier l'écriture on posera

$$\begin{cases} \phi_i = \phi \circ t_i \\ \mu_i = \mu \circ \phi_i = \mu \circ \phi \circ t_i \end{cases} \tag{7}$$

ϕ_i et μ_i sont des applications de $\mathbb{R}^m$ dans respectivement $\mathbb{R}^{k*}$ et $\mathbb{R}$

Ainsi nous avons

$$L'_n(\omega,\theta,\xi) = \alpha(\xi) \sum_{i=1}^{n} [<\phi'_i(\theta),\omega_i> - \mu'_i(\theta)]$$

on peut écrire

$$\mu'_i(\theta) = \mu'(\gamma_i) \circ \phi'_i(\theta) = <\mu'(\gamma_i),\phi'_i(\theta)> = <\phi'_i(\theta), \mu'(\gamma_i)>$$

et donc

$$L'_n(\omega,\theta,\xi) = \alpha(\xi) \sum_{i=1}^{n} <\phi'_i(\theta),\omega_i-\mu'(\gamma_i)> \tag{8}$$

où $\phi'_i(\theta)$ est considéré comme un élément de $\mathcal{L}(\mathbb{R}^m,\mathbb{R}^{k*})$

$\mu'(\gamma_i)$ comme un élément de $\mathcal{L}(\mathbb{R}^{k*},\mathbb{R})$

on identifie $\mathcal{L}(\mathbb{R}^{k*},\mathbb{R})$ avec $\mathbb{R}^k$

$L'_n(\omega,\theta,\xi)$ est donc considéré comme un élément de $\mathbb{R}^{m*}$.

$$\forall z \in \mathbb{R}^m \; [L'_n(\omega,\theta,\xi)](z) = \alpha(\xi) \sum_{i=1}^{n} <[\phi'_i(\theta)](z), \omega_i-\mu'(\gamma_i)>.$$

De même on a

$$L''_n(\omega,\theta,\xi) = \alpha(\xi) [\sum_{i=1}^{n} <\phi''_i(\theta),\omega_i-\mu'(\gamma_i)> - \sum_{i=1}^{n} <\mu''(\gamma_i) \circ \phi'_i(\theta),\phi'_i(\theta)>]$$

avec $\mu''(\gamma_i) \in \mathcal{L}^2(\mathbb{R}^{k*},\mathbb{R})$

on identifie $\mathcal{L}^2(\mathbb{R}^{k*},\mathbb{R})$ avec $\mathcal{L}(\mathbb{R}^{k*},\mathbb{R}^k)$

$\phi''_i(\theta) \in \mathcal{L}^2(\mathbb{R}^m,\mathbb{R}^{k*})$

donc $L''_n(\omega,\theta,\xi) \in \mathcal{L}^2(\mathbb{R}^m,\mathbb{R})$

$\forall (z_1,z_2) \in \mathbb{R}^m \times \mathbb{R}^m$ on a $L''_n(\omega,\theta,\xi)(z_1,z_2) =$

$$\alpha(\xi) \sum_{i=1}^{n} [<[\phi''_i(\theta)](z_1,z_2),\omega_i-\mu'(\gamma_i)> - [\mu''(\gamma_i)]([\phi'_i(\theta)](z_1),[\phi'_i(\theta)](z_2))]$$

dans la suite on pose $<\mu''(\gamma_i) \circ \phi_i'(\theta), \phi_i'(\theta)> = \Gamma_i(\theta)$

où $\Gamma_i(\theta) \in \mathcal{L}^2(\mathbb{R}^m,\mathbb{R})$

on a alors :

$$L_n''(\omega,\theta,\xi) = \alpha(\xi) \sum_{i=1}^{n} [<\phi_i''(\theta),\omega_i-\mu'(\gamma_i)>-\Gamma_i(\theta)] \qquad (9)$$

On a également

$$L_n'''(\omega,\theta,\xi) = \alpha(\xi) \sum_{i=1}^{n} [<\phi_i'''(\theta),\omega_i-\mu'(\gamma_i)>-<\phi_i''(\theta),\mu''(\gamma_i) \circ \phi_i'(\theta)> - \Gamma_i'(\theta)]$$

$$\left.\begin{array}{ll} \text{avec } \phi_i'''(\theta) \in \mathcal{L}^3(\mathbb{R}^m,\mathbb{R}^{k*}) \\ \Gamma_i'(\theta) \in \mathcal{L}^3(\mathbb{R}^m,\mathbb{R}) \end{array}\right\} \Longrightarrow L_n'''(\omega,\theta,\xi) \in \mathcal{L}^3(\mathbb{R}^m,\mathbb{R})$$

on pose $\Pi_i(\theta) = \Gamma_i'(\theta) + <\phi_i''(\theta),\mu''(\gamma_i) \circ \phi_i'(\theta)>$

on a donc

$$L_n'''(\omega,\theta,\xi) = \alpha(\xi) \sum_{i=1}^{n} [<\phi_i'''(\theta), \omega_i-\mu'(\gamma_i)> - \Pi_i(\theta)] \qquad (10)$$

En vue de simplifier les écritures on note

$\mu_i'(\phi_i(\theta)) = \overline{\mu_i'} \qquad \phi_i'(\overline{\theta}) = \overline{\phi_i'}$

ainsi de suite.

On note aussi

$L_n(.,\overline{\theta},\xi) = \overline{L_n} \qquad L_n'(.,\overline{\theta},\xi) = \overline{L_n'}$

ainsi de suite.

II.3.3 - Hypothèse sur la moyenne de Ω_i

Conformément à ce que l'on a pour le modèle linéaire généralisé unidimensionnel on suppose que :

$\forall\theta,\xi \in \Theta \times \Xi \qquad$ on a $E_{\theta,\xi}(\Omega_i) = \mu'(\gamma_i) = \mu'(\phi(t_i\theta))$

ainsi θ est le seul paramètre qui intervient dans l'expression de la moyenne de Ω_i .

II.3.4 - Définition d'un produit scalaire sur $\mathbb{R}^m$ et $\mathbb{R}^{m*}$

On suppose que pour tout θ appartenant à Θ la forme bilinéaire symétrique et définie positive $\sum_{i=1}^{n} \frac{\Gamma_i(\theta)}{n}$ admet une limite notée $\Gamma(\theta)$.

$\Gamma(\theta)$ est elle même alors une forme bilinéaire et symétrique.

On suppose également que $\Gamma(\theta)$ est définie positive. On suppose que Γ est une application continue de θ dans $\mathcal{L}^2(\mathbb{R}^m,\mathbb{R})$.

θ étant un élément donné de $\mathbb{R}^m$ on munit $\mathbb{R}^m$ du produit scalaire noté $<\cdot,\cdot>_\theta$ lequel est défini par

$$\forall\theta_1,\theta_2 \in \mathbb{R}^m \; <\theta_1,\theta_2>_\theta = [\Gamma(\theta)]\,(\theta_1,\theta_2).$$

On note également $\Gamma(\theta)$ la bijection de $\mathbb{R}^m$ sur son dual $\mathbb{R}^{m*}$ définie par

$$\theta_1^* = [\Gamma(\theta)](\theta_1) \iff \forall\theta' \in \mathbb{R}^m \quad \theta_1^*(\theta') = <\theta_1,\theta'>_\theta$$

On peut aussi munir $\mathbb{R}^{m*}$ d'un produit scalaire que l'on note aussi $<.,.>_\theta$ et défini par

$$\forall\theta_1^*,\theta_2^* \in \mathbb{R}^{m*} \quad <\theta_1^*,\theta_2^*>_\theta = <[\Gamma(\theta)]^{-1}(\theta_1^*), \; [\Gamma(\theta)]^{-1}(\theta_2^*)>_\theta$$

$\Gamma(\theta)$ est ainsi une isométrie de $\mathbb{R}^m$ sur $\mathbb{R}^{m*}$

Dans la suite on utilise surtout Γ au point $\bar\theta$ aussi pour simplifier l'écriture on note

$$\Gamma(\bar\theta) \text{ par } \bar\Gamma$$
$$\Gamma_i(\bar\theta) \text{ par } \bar\Gamma_i$$
$$<.,.>_{\bar\theta} \text{ par } <.,.>$$

de même $<\theta,\theta>_{\bar\theta} = \|\theta\|^2_{\bar\theta}$ par $\|\theta\|^2$

II.3.5 - Conditions de régularité sur les dérivées de $L_n(.,\theta,\xi)$

Si $B(\bar\theta,\rho)$ est définie de la manière suivante

$$B(\bar\theta,\rho) = \{\theta, \; \|\theta - \bar\theta\| \leq \rho\}$$

On suppose que pour tout nombre ρ positif les applications $\phi_i',\mu_i',\phi_i'',\mu_i'',\phi_i'''$, π_i sont bornées sur $B(\bar\theta,\rho)$

II.3.6 - Condition concernant $\Omega_i-\mu_i'$

On suppose qu'il existe un nombre strictement positif δ ainsi qu'un voisinage de $\bar\theta$ que l'on appellera $V(\bar\theta)$ tels que :

$$\sup_{i\in\mathbb{N},\theta\in V(\bar\theta)} E_{\theta,\xi}[\|\Omega_i-\mu_i'\|^{2+\delta}] < d < +\infty$$

II.4 - Estimateurs de θ

Dans le modèle linéaire généralisé la détermination de l'estimateur du maximum de vraisemblance pouvant poser quelques difficultés pratiques on considère la classe d'estimateurs $\hat{\theta}_n$ satisfaisant

$$\begin{cases} \frac{1}{\sqrt{n}} L'_n(.,\hat{\theta}_n) \to 0 \quad \text{en } P_n \text{ Probabilité} \\ \mathcal{L}(\sqrt{n}(\hat{\theta}_n-\bar{\theta})/P_n) \text{ a une limite} \end{cases} \tag{11}$$

Pour montrer que cette classe est non vide on montre qu'il existe une solution des équations de vraisemblance satisfaisant la seconde condition, pour cela on peut étudier la limite en loi de L'_n à la fois pour la valeur $\bar{\theta}$ et pour la valeur θ_n (on rappelle $\theta_n = \bar{\theta} + \frac{h_n}{\sqrt{n}}$) du paramètre θ , la seconde se déduisant de la première une fois montrée la contiguïté des suites $\{P_n\}$ et $\{P'_n\}$.

II.4.1 - Limite de $\mathcal{L}(\frac{L'_n}{\sqrt{n}} / P_n)$

Proposition II_1 : *Sous les conditions II_3*

$$\mathcal{L}(\frac{L'_n}{\sqrt{n}} / P_n) \to N(0,\alpha(\xi)\ \bar{\Gamma})$$

loi normale de moyenne nulle de matrice de variances et covariances $\alpha(\xi)\bar{\Gamma}$.

II.4.2 - Contiguïté de $\{P_n\}$ et $\{P'_n\}$

Définition : *Soit $(\Omega,\mathcal{A}_n)$ une suite d'espace probabilisable $\{P_n\}$ et $\{P'_n\}$ étant 2 suites de mesures de probabilités sur $(\Omega,\mathcal{A}_n)$; ces deux suites sont dites contigües si pour toute suite $\{T_n\}$ de variables aléatoires définies sur $(\Omega,\mathcal{A}_n)$ les 2 propriétés suivantes sont équivalentes*

- *T_n converge vers 0 en P_n probabilité*
- *T_n converge vers 0 en P'_n probabilité.*

Proposition II_2 : *Sous la condition II_3*

Les deux suites $\{P_n\}$ et $\{P'_n\}$ sont contigües.

En conséquence si une suite de variables aléatoires T_n converge vers 0 en P_n-probabilité ou en P'_n-probabilité on dira simplement que T_n converge en probabilité vers 0 et on notera $T_n \xrightarrow{P} 0$.

Remarque : Si $\{P_n\}$ et $\{P'_n\}$ sont contigües, T_n converge vers une constante a en P_n-probabilité est équivalent à T_n converge vers a en P'_n-probabilité ; en effet $T_n \xrightarrow{P_n} a$ est équivalent à $T_n - a \xrightarrow{P_n} 0$.
Pour une étude de la contiguïté on se réfèrera à Roussas (1972).

II.4.3 - Limite de $\mathcal{L}\left(\frac{\overline{L}'_n}{\sqrt{n}} / P'_n\right)$

Proposition II_3 : *Sous les conditions* II_3

$$\mathcal{L}\left(\frac{\overline{L}'_n}{\sqrt{n}} / P'_n\right) \to N(\alpha(\xi)\ \overline{\Gamma}(h), \alpha(\xi)\overline{\Gamma})$$

II.4.4 - Comportement asymptotique de l'estimateur du maximum de vraisemblance

Proposition II_4 : *Il existe une solution de l'équation*

$$L'_n(.,\theta,\xi) = 0 \qquad (12)$$

qui converge vers $\overline{\theta}$ *en probabilité ; on la note* $\hat{\theta}_n$.

On note $\Gamma_n = \Gamma(\hat{\theta}_n)$

on a $\Gamma_n \xrightarrow{P} \overline{\Gamma}$

En effet $\hat{\theta}_n \xrightarrow{P} \theta$ et l'application Γ est continue de θ dans $\mathcal{L}^2(\mathbb{R}^m, \mathbb{R})$

Proposition II_5 : *Sous les conditions* II_3 *et si* $\hat{\theta}_n$ *est la solution convergente de* (12), *on a les résultats suivants.*

$$\mathcal{L}(\sqrt{n}(\hat{\theta}_n - \overline{\theta})/P_n) \longrightarrow N(0; \frac{1}{\alpha(\xi)}\ \overline{\Gamma}^{-1})$$

et

$$\mathcal{L}(\sqrt{n}(\hat{\theta}_n - \overline{\theta})/P'_n) \longrightarrow N(h; \frac{1}{\alpha(\xi)}\ \overline{\Gamma}^{-1})$$

II.5 - Structure gaussienne tangente

$\hat{\theta}_n$ désignant une statistique quelconque satisfaisant la condition (II) on peut montrer que $\hat{\theta}_n$ est très voisin de l'estimateur de θ dans une structure gaussienne tangente .

Soit Y_i la variable aléatoire à valeur dans $\mathbb{R}^m$ définie de la manière suivante :

$$Y_i(\omega_i) = \bar{\Gamma}^{-1} < \bar{\phi}_i' \; , \; \omega_i - \bar{\mu}_i' > + \bar{\theta} \qquad (13)$$

où $\left.\begin{array}{l} \bar{\phi}_i' \in \mathcal{L}(\mathbf{R}^m, \mathbf{R}^k) \\ \omega_i - \bar{\mu}_i' \in \mathbf{R}^k \end{array}\right\} \Longrightarrow <\phi_i', \omega_i - \bar{\mu}_i'> \in \mathbf{R}^{m*}$

et posons

$$\tilde{\theta}_n = \sum_{i=1}^{n} \frac{Y_i}{n} \qquad (14)$$

<u>Proposition II_6</u> : *Sous les conditions II_3*

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) \xrightarrow{P} 0$$

et

$$\mathcal{L}(\sqrt{n}(\tilde{\theta}_n - \bar{\theta})/P_n') \longrightarrow N(h; \frac{1}{\alpha(\xi)} \bar{\Gamma}^{-1})$$

On peut interpréter ce résultat de la manière suivante :
Si Y_i était une variable aléatoire suivant une loi $N(\theta; \frac{1}{\alpha(\xi)} \quad \bar{\Gamma}^{-1})$, on disposerait d'un échantillon $y = (y_1, y_2, \ldots, y_n)$ avec $y_i = Y_i(\omega_i)$ de telle façon que l'on peut se placer dans la structure statistique suivante :

$$\{ \mathbf{R}^m, \mathcal{B}_{\mathbf{R}^m} \; , \; N_m [\theta, \frac{1}{\alpha(\xi)} \quad \bar{\Gamma}^{-1}] \quad \theta \in \mathbf{R}^m, \alpha(\xi) \in \mathbf{R}_+^* \}^n \qquad (15)$$

Le logarithme de la vraisemblance serait alors :

$$\tilde{L}_n(y, \theta, \xi) = \text{cte} - \frac{1}{2} \sum_{i=1}^{n} \alpha(\xi) \; \bar{\Gamma}(y_i - \theta, y_i - \theta)$$

l'estimateur du maximum de vraisemblance serait alors

$$\sum_{i=1}^{n} \frac{Y_i}{n} = \tilde{\theta}_n$$

La proposition II_6 justifie a posteriori l'appellation de structure gaussienne tangente.

Pour plus de précisions sur la notion de structure gaussienne tangente on renvoie le lecteur à Lecam (1969), Roussas (1972), Mathieu (1978) :
Les démonstrations des propositions II_1, II_2, II_3, II_4 , II_5 , II_6 peuvent être trouvées dans Mathieu (1978, 1981).
La plupart de ces résultats ont été démontrés aussi, dans un cadre légèrement différent, dans Philippou et Roussas (1973, 1975).

CHAPITRE III

TESTS DU $\bar{\chi}^2$ DANS LE MODÈLE LINÉAIRE GÉNÉRALISÉ

- o - o - o - o -

II.1 - Introduction

Soit la structure statistique précédemment considérée

$$\bigotimes_{i=1}^{n} \left\{ \mathbb{R}^k , \mathcal{B}_{\mathbb{R}^k} , f_i(.,\theta,\xi) \quad \theta \in \Theta \quad \xi \in \Xi \right\}$$

où

$f_i(u,\theta,\xi)$ a la forme donnée au chapitre II (4)

Soit les hypothèses

$$H_0 : \theta = 0 \Longleftrightarrow \theta_i = 0 \qquad \forall\ i=1,\ldots,m$$

$$H : \theta_i \geq 0 \qquad i = 1,\ldots,m$$

$$H_1 : H \setminus H_0$$

On s'intéresse dans ce chapitre au test de l'hypothèse H_0 contre l'hypothèse H_1.

C'est par abus de langage que l'on dit test de H_0 contre H_1, car il s'agit en fait de $H_0 \times \Xi$ contre $H_1 \times \Xi$; c'est-à-dire que aucune des hypothèses évoquées n'impose de contrainte à ξ , ξ joue le rôle de paramètre fantôme. C'est donc aussi par abus de langage que l'on dira que l'hypothèse H_0 est simple. Ce travail peut être considéré comme une généralisation de ce qui a été fait au chapitre I au cas où les lois ne sont pas normales. En 1978 Robertson et Wegman ont présenté une extension des tests d'hypothèses d'isotonie dans le cas où l'on a affaire à la famille exponentielle. Notre travail découle de la même idée de généralisation, seulement nous nous plaçons dans le cadre du modèle linéaire généralisé multidimensionnel et au lieu de considérer les hypothèses d'isotonie on considère les mêmes hypothèses que Kudô (1963) Kudô et Fujisawa (1965) Kudô et Choï (1975). De manière évidente, dans le modèle linéaire généralisé, de nombreux types d'hypothèses, imposent des contraintes d'inégalités aux composantes du paramètre θ, peuvent se ramener, par un simple changement de variables, aux hypothèses que l'on considère ; en effet θ intervient toujours dans le modèle sous la forme $t_i\,\theta$, où les t_i sont des applications linéaires connues, donc pour changer la forme des hypothèses il suffit de changer de t_i.

On s'intéressera aussi dans ce chapitre au test de l'hypothèse H contre l'hypothèse $\mathbb{R}^m \setminus H$.

L'idée du travail est classique, en statistique asymptotique, il s'agit d'obtenir des statistiques de test de H_0 contre H_1 dont la distribution asymptotique se déduira de la normalité asymptotique de certains estimateurs. Dans ce qui nous intéresse on cherche des statistiques dont les distributions asymptotiques seront des lois de $\bar{\chi}^2$. Robertson et Wegman (1978) ont montré que le test du rapport de vraisemblance des hypothèses d'isotonie dans une famille exponentielle était fait à partir d'une distribution asymptotique de $\bar{\chi}^2$ correspondant à un ordre et à des poids déterminés. Par analogie la loi de $\bar{\chi}^2$ que nous obtenons dépend d'une matrice à déterminer.

III.2 - Estimation de θ sous contrainte d'appartenance à un cône convexe polyédrique.

Afin de pouvoir établir la statistique du test du rapport de vraisemblance de l'hypothèse H_0 contre l'hypothèse H_1 il est indispensable de disposer d'une estimation de θ sous la contrainte H. H étant un cône convexe polyhédrique, La solution des équations de vraisemblance sous la contrainte H peut s'avérer très délicate et on peut dans certains cas être amené à ne connaître que des valeurs approchées de l'estimateur du maximum de vraisemblance ; on se demande donc si la projection de $\hat{\theta}_n(\omega)$ sur H, où $\hat{\theta}_n$ est l'un quelconque des estimateurs de satisfaisant la condition II(11) ne constitue pas une approximation satisfaisante de l'estimateur recherché permettant de construire un test de H_0 contre H_1 ayant les mêmes propriétés asymptotiques que le test du rapport de vraisemblance. L'hypothèse nulle étant simple, dans ce chapitre, on considère $\bar{\theta} = 0$. $\hat{\theta}_n$ étant une statistique satisfaisant II(11) on note $\hat{\theta}_{H,n}$ la projection de $\hat{\theta}_n$ sur H au sens du produit scalaire défini par $\bar{\Gamma}$. De même que pour l'estimateur $\hat{\theta}_n$ on veut montrer que $\hat{\theta}_{H,n}$ est proche d'une statistique que l'on peut définir dans la structure gaussienne tangente ; on introduit donc l'estimateur $\hat{\theta}_{H,n}$ de θ sous la contrainte H dans la structure II(15). $\tilde{\theta}_{H,n}$ est la projection de $\tilde{\theta}_n$ sur H au sens du produit scalaire défini par $\bar{\Gamma}$. On voit donc ici aussi qu'il est nécessaire de posséder une méthode de projection sur un cône convexe polyhédrique.

Proposition III.1 : *Sous les conditions II.3*

$$\sqrt{n}(\tilde{\theta}_{H,n} - \hat{\theta}_{H,n}) \xrightarrow{P} 0$$

Démonstration

Cela est dû au fait que l'on ait, d'après la proposition II.6, $\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) \xrightarrow{P} 0$ et au fait que la projection sur un convexe fermé est une application contractante i.e. $\|\hat{\theta}_{H,n} - \tilde{\theta}_{H,n}\| \leq \|\hat{\theta}_n - \tilde{\theta}_n\|$.

Si nous avons pu jusqu'à présent négliger le paramètre ξ , afin de construire des statistiques de test on est contraint de posséder une estimation de $\alpha(\xi)$ on fait la supposition suivante

[I.3 - Conditions sur ξ

Il existe un estimateur convergent de ξ , on note cet estimateur $\hat{\xi}$, de plus on suppose que la fonction α de Ξ dans $\mathbb{R}$ est continue et qu'il en est de même pour la fonction $\psi(u,\cdot)$.
Ainsi on dispose de $\alpha(\hat{\xi})$ estimateur convergent de $\alpha(\xi)$.
La connaissance d'un estimateur convergent de $\alpha(\xi)$ est souvent assurée par le fait que $\alpha(\xi)$ est connu, cette condition est satisfaite aussi dans le cas gaussien avec σ^2 inconnue.

On pose $T_{\tilde{H}_{o,1}} = n\,\alpha(\xi)\,\|\tilde{\theta}_{H,n}\|^2$

I.4 - Distribution asymptotique de $T_{H_{o,1}}$

On utilise le lemme suivant que l'on trouvera dans Billingsley (1968) par exemple

Lemme : *$X_n \xrightarrow{\mathcal{L}} X$. X_n et X sont des variables aléatoires à valeur dans un espace topologique si h est une application continue sur cet espace $h \circ X_n = h(X_n)$ et $h \circ X = h(X)$ sont des variables aléatoires et $h(X_n) \xrightarrow{\mathcal{L}} h(X)$.*

Proposition III.2 - *Si $X_n \longrightarrow N_m(0,\Sigma)$*

Si $X_{H,n}$ projection de X_n sur H (le quadrant positif de $\mathbb{R}^m$) au sens du produit scalaire Σ^{-1}

alors $\|X_{H,n}\|^2_{\Sigma^{-1}} \longrightarrow \bar{\chi}^2_{\Sigma}$

de même si $X_n \longrightarrow N_m(h,\Sigma)$

alors $\|X_{H,n}\|^2_{\Sigma^{-1}} \longrightarrow \bar{\chi}^2_{\Sigma,h}$ (loi de $\bar{\chi}^2$ décentrée)

Cela découle immédiatement du lemme, de la proposition I_5 et de la définition des lois de $\bar{\chi}^2$ décentrées.

II.4.1 - Distribution de $T_{\tilde{H}_{o,1}}$ sous P_n

On sait que $\mathcal{L}(\sqrt{n}\,\tilde{\theta}_n/P_n) \longrightarrow N(0, \frac{1}{\alpha(\xi)}\,\bar{\Gamma}^{-1})$

d'après la proposition III.2 et comme $\tilde{\theta}_{H,n}$ est la projection de $\tilde{\theta}_n$ sur H on a

$$\mathcal{L}(n\,\alpha(\xi)\,\bar{\Gamma}(\tilde{\theta}_{H,n}, \tilde{\theta}_{H,n}) / P_n) \longrightarrow \bar{\chi}^2_{\bar{\Gamma}^{-1}} .$$

autrement dit :

$$\mathcal{L}(T_{\tilde{H}_{o,1}} / P_n) \longrightarrow \bar{\chi}^2_{\bar{\Gamma}^{-1}} .$$

III.4.2. Distribution de $T_{H_{0,1}}$ sous P'_n

On a $\mathcal{L}(\sqrt{n}\ \tilde{\theta}_n/P'_n) \longrightarrow N(h, \frac{1}{\alpha(\xi)}\ \bar{\Gamma}^{-1})$

d'après la proposition III.2 on a donc

$$\mathcal{L}(n\ \ \alpha(\xi)\ \bar{\Gamma}(\tilde{\theta}_{H,n}\ ,\ \tilde{\theta}_{H,n})\ /\ P'_n) \longrightarrow \bar{\chi}^2_{\bar{\Gamma}^{-1},h}$$

Soit

$$\mathcal{L}(T_{\tilde{H}_{0,1}}\ /\ P'_n) \longrightarrow \bar{\chi}^2_{\bar{\Gamma}^{-1},h}$$

$\bar{\chi}^2_{\bar{\Gamma}^{-1},h}$ désigne la loi de $\bar{\chi}^2$ décentrée dépendant de la matrice $\bar{\Gamma}^{-1}$ et du paramètre de décentrage h.

III.5. Un test de H_0 contre H_1

Pour faire le test de H_0 contre H_1 on peut procéder de la façon suivante :

On définit la statistique $T_{1,H_{0,1}}$ par

$$T_{1,H_{0,1}} = n\ \alpha(\hat{\xi})\ \|\hat{\theta}_{H,n}\|^2$$

et on définit un test de la manière suivante

rejet de H_0 pour $T_{1,H_{0,1}} \geq \ell$

Soit $\phi_{1,H_{0,1}} = 1_{\{T_{1,H_{0,1}} \geq \ell\}}$

Proposition III.3 : *Sous les conditions II.3 et III.3 on a le résultat suivant*

$$T_{1,H_{0,1}} - T_{\tilde{H}_{0,1}} \xrightarrow{P} 0\ .$$

Démonstration : D'après la condition III.3 $\hat{\xi} \xrightarrow{P} \xi$

$$\alpha(\hat{\xi}) \xrightarrow{P} \alpha(\xi)$$

d'après la proposition III.1 on a $\sqrt{n}(\hat{\theta}_{H,n} - \tilde{\theta}_{H,n}) \xrightarrow{P} 0$

$$T_{1,H_{0,1}} - T_{H_{0,1}} = n\ \alpha(\hat{\xi})\ \|\hat{\theta}_{H,n}\|^2 - n\ \alpha(\xi)\ \|\theta_{H,n}\|^2 =$$

$$= n\ \alpha(\hat{\xi})(\ \|\hat{\theta}_{H,n}\|^2 - \|\tilde{\theta}_{H,n}\|^2\ ) + n\ \|\tilde{\theta}_{H,n}\|^2\ \ (\alpha(\hat{\xi}) - \alpha(\xi))$$

$\sqrt{n}\ \tilde{\theta}_n$ a une limite en loi d'où $\sqrt{n}\,\tilde{\theta}_{H,n}$ a une limite en loi

donc $n\,\|\tilde{\theta}_{H,n}\|^2$ a une limite en loi

donc $n\,\|\tilde{\theta}_{H,n}\|^2\ (\alpha(\hat{\xi}) - \alpha(\xi)) \xrightarrow{P} 0$

on a aussi $n(\|\hat{\theta}_{H,n}\|^2 - \|\tilde{\theta}_{H,n}\|^2) \xrightarrow{P} 0$

d'où le résultat.

Remarques

Pour faire un test de niveau asymptotique α , on utilisera le fait que sous H_0, $T_{H_{o,1}}$ tend vers une loi de $\bar{\chi}^2_{\bar{\Gamma}-1}$ la valeur ℓ_α pour laquelle on rejettera H_0 sera défini à partir de cette loi. De même pour calculer la "puissance" asymptotique on utilisera la loi $\bar{\chi}^2_{\bar{\Gamma}-1,h}$.

Il est important de noter que la loi $\bar{\chi}^2_{\bar{\Gamma}-1}$ dépend de la vraie valeur du paramètre θ , comme l'hypothèse H_0 est "simple" (à ξ près) il n'y a pas de problème.

II.6 - Test du rapport de vraisemblance de H_0 contre H_1.

On définit la statistique $T_{2,H_{o,1}}$ de la manière suivante :

$$T_{2,H_{o,1}} = 2[L_n(.,\hat{\theta}_{H,n},\hat{\xi}) - L_n(.\,,\hat{\theta}_{H_{o,n}},\hat{\xi})]$$
$$= 2[L_n(.,\hat{\theta}_{H,n},\hat{\xi}) - L_n(.,0,\hat{\xi})]$$

et on définit le test du rapport de vraisemblance par

$$\text{rejet de } H_0 \text{ pour } T_{2,H_{o,1}} \geq \ell$$

$$\text{soit } \phi_{2,H_{o,1}} = 1_{\{T_{2,H_{o,1}} \geq \ell\}}$$

Proposition III.4 : *Sous les conditions II.3 et III.3*
$T_{2,H_{0,1}} - T_{\tilde{H}_{0,1}} \xrightarrow{P} 0$.

Démonstration :

$$T_{2,H_{o,1}} = 2[L_n(.,\hat\theta_{H,n},\hat\xi) - L_n(.,\hat\theta_{H_{o,n}},\hat\xi)]$$
$$= 2[L_n(.,\hat\theta_n,\hat\xi) - L_n(.,\hat\theta_{H_{o,n}},\hat\xi)]$$
$$-2[L_n(.,\hat\theta_n,\hat\xi) - L_n(.,\hat\theta_{H,n},\hat\xi)]$$

de même

$$T_{\tilde H_{o,1}} = n\,\alpha(\xi)\,\|\tilde\theta_{H,n}\|^2 = n\alpha(\xi)(\|\tilde\theta_n\|^2 - \|\tilde\theta_n - \tilde\theta_{H,n}\|^2)$$

$$T_{2,H_{o,1}} - T_{\tilde H_{o,1}} = 2[L_n(.,\hat\theta_n,\hat\xi) - L_n(.,\hat\theta_{H_{o,n}},\hat\xi)] - n\alpha(\xi)\,\|\tilde\theta_n\|^2$$
$$-2[L_n(.,\hat\theta_n,\hat\xi) - L_n(.,\hat\theta_{H,n},\xi)] + n\alpha(\xi)\,\|\tilde\theta_n - \tilde\theta_{H,n}\|^2$$

Dans Mathieu (1981) proposition 9 il est montré que

$$2[L_n(.,\hat\theta_n,\hat\xi) - L_n(.,\hat\theta_{H_{o,n}},\hat\xi)] - n\,\alpha(\xi)\,\|\tilde\theta_n\|^2 \xrightarrow{P} 0$$

On doit montrer que $2\,L_n(.,\hat\theta_n,\hat\xi) - L_n(.,\hat\theta_{H,n},\hat\xi) - n\alpha(\xi)\,\|\tilde\theta_n - \tilde\theta_{H,n}\|^2 \xrightarrow{P} 0$

$$L_n(.,\hat\theta_{H,n},\xi) - L_n(.,\hat\theta_n,\hat\xi) - [L'_n(.,\hat\theta_n,\hat\xi)]\,(\hat\theta_{H,n} - \hat\theta_n)$$
$$- \frac{1}{2}[L''_n(.,\hat\theta_n,\hat\xi)]\,(\hat\theta_{H,n} - \hat\theta_n,\hat\theta_{H,n} - \hat\theta_n) \xrightarrow{P} 0$$

$$\frac{1}{\sqrt n}\,L'_n(.,\hat\theta_n,\hat\xi) \xrightarrow{P} 0$$

$\sqrt n\,\hat\theta_n$ a une limite en loi .

$\sqrt n(\hat\theta_n - \hat\theta_{H,n})$ est la projection de $\hat\theta_n$ sur le cone polaire de H d'où $\sqrt n(\hat\theta_n - \hat\theta_{H,n})$ a une limite en loi

d'où $\frac{1}{\sqrt n}\,L'_n(.,\hat\theta_n,\hat\xi).(\sqrt n\,(\hat\theta_{H,n} - \hat\theta_n)) \xrightarrow{P} 0$

$$\|\frac{1}{n}\,L''_n(.,\hat\theta_n,\hat\xi) - \bar\Gamma\| \xrightarrow{P} 0$$

donc $2[L_n(.,\hat\theta_{H,n},\hat\xi) - L_n(.,\hat\theta_n,\hat\xi)] + n\alpha(\hat\xi)\,\|\hat\theta_{H,n} - \hat\theta_n\|^2 \xrightarrow{P} 0$

on a $n\alpha(\hat\xi)\,\|\hat\theta_{H,n} - \hat\theta_n\|^2 - n\alpha(\xi)\,\|\tilde\theta_{H,n} - \tilde\theta_n\|^2 \xrightarrow{P} 0$

d'où le résultat.

Remarque : La démonstration du fait que

$$2[L_n(.,\hat{\theta}_n,\xi) - L_n(.,\hat{\theta}_{H,n},\hat{\xi})] - n\alpha(\hat{\xi})\ \|\hat{\theta}_n - \hat{\theta}_{H,n}\|^2 \xrightarrow{P} 0$$

est analogue à la démonstration de la proposition 9 de Mathieu (1981) on utilise en effet le fait que $\sqrt{n}(\hat{\theta}_{H,n} - \hat{\theta}_n)$ ait une limite en loi dans Mathieu H désigne un sous-espace de $\mathbb{R}^m$ dans notre cas H désigne un cône convexe polyédrique ; cela n'ajoute aucune difficulté.

Remarque : On aurait pu s'intéresser au test de H_o contre H_1 dans le cas où

$$H_o : \theta = \bar{\theta} \iff \theta_i = \bar{\theta}_i \quad \forall\ i=1,\dots,m$$

$$H : \theta_i \geq \bar{\theta}_i \quad \forall\ i=1,\dots,m$$

$$H_1 = H \setminus H_o$$

dans ce cas les résultats sont quasiment identiques, il suffit de considérer

$$T_{\tilde{H}_{o,1}} = n\alpha(\xi)\ \|\tilde{\theta}_{H,n} - \bar{\theta}\|^2$$

$$T_{1,H_{o,1}} = n\alpha(\hat{\xi})\ \|\hat{\theta}_{H,n} - \bar{\theta}\|^2$$

$$T_{2,H_{o,1}} = 2[L_n(.,\hat{\theta}_{H,n},\xi) - L_n(.,\bar{\theta},\hat{\xi})]$$

Remarque : Les 2 tests $\phi_{1\ H_{o,1}}$ et $\phi_{2,H_{o,1}}$ sont asymptotiquement équivalents, c'est à dire que l'évènement $\{|\phi_{1,H_{o,1}} - \phi_{2,H_{o,1}}| \neq 0\}$ est asymptotiquement de probabilité nulle pour P_n et P'_n.

En effet, soit $\varepsilon > 0$

$$\phi_{1,H_{o,1}} = 1_{\{T_{1,H_{o,1}} \geq \ell\}}$$

$$\phi_{2,H_{o,1}} = 1_{\{T_{2,H_{o,1}} \geq \ell\}}$$

$$|\phi_{1,H_{o,1}} - \phi_{2,H_{o,1}}| \leq 1_{|T_{1,H_{o,1}} - T_{2,H_{o,1}}| \geq \varepsilon} + 1_{\ell < T_{1,H_{o,1}} < \ell+\varepsilon} + 1_{\ell < T_{2,H_{o,1}} < \ell+\varepsilon}$$

Soit $\gamma > 0$

si n est suffisamment grand et ε suffisamment petit comme $T_{1,H_{o,1}} - T_{\tilde{H}_{o,1}} \xrightarrow{P} 0$

$T_{2,H_{o,1}} - T_{\tilde{H}_{o,1}} \xrightarrow{P} 0$

on a $T_{1,H_{o,1}} - T_{2,H_{o,1}} \xrightarrow{P} 0$

et $P_n(|T_{1,H_{o,1}} - T_{2,H_{o,1}}| \geq \varepsilon] < \frac{\gamma}{3}$

on a $P_n\ [\ell < T_{1,H_{o,1}} < \ell + \varepsilon] < \frac{\gamma}{3}$

$P_n [\ \ell < T_{2,H_{o,1}} < \ell + \varepsilon] < \frac{\gamma}{3}$

d'où $\mathbb{E}[|\phi_{1,H_{o,1}} - \phi_{2,H_{o,1}}|] < \gamma$

c'est-à-dire

$$P_n [\ |\phi_{1,H_{o,1}} - \phi_{2,H_{o,1}}| \neq 0\ ] < \gamma$$

d'où le résultat.

III.7 - Tests de H contre $\mathbb{R}^m \setminus H$

On peut, comme on l'a fait dans le cas des lois normales au chapitre I paragraphe 5, au test de l'hypothèse H

$$H : \theta_i \geq 0 \qquad \forall\ i=1,\dots,m$$

contre l'hypothèse $\mathbb{R}^k \setminus H$.

On pose donc

$$T_{\tilde{H}} = n\alpha(\xi)\ \|\tilde{\theta}_n - \tilde{\theta}_{H,n}\|^2 \quad .$$

On va comme dans le cas des tests de H_o contre H_1 construire des statistiques de tests dont les distributions asymptotiques seront égales à celle de $T_{\tilde{H}}$.
On a donc besoin de connaître la distibution asymptotique de $T_{\tilde{H}}$.
On établit un résultat qui va nous permettre de calculer la distribution asymptotique de $T_{\tilde{H}}$.

Proposition III.5 : *Si* $X_n \longrightarrow N_m(0,\Sigma)$

$X_{H,n}$ projection de X_n sur H au sens du produit scalaire induit par Σ^{-1}

$\|X_n - X_{H,n}\|^2_{\Sigma^{-1}} \longrightarrow \bar{\chi}^2_{(\Sigma^0)}$ *Loi de $\bar{\chi}^2$ définie à la proposition I.8.*

Cela découle du lemme du paragraphe II.4.

Corollaire : $(T_{\tilde{H}}/P_n) \longrightarrow \bar{\chi}^2_{(\bar{\Gamma}^{-1},0)}$

En effet $\mathcal{L}(\sqrt{n}\,\tilde{\theta}_n/P_n) \longrightarrow N(0, \frac{1}{\alpha(\xi)}\bar{\Gamma}^{-1})$

$\tilde{\theta}_{H,n}$ étant la projection de $\tilde{\theta}_n$ sur le cône convexe H d'après la proposition précédente on a

$$\mathcal{L}(n\,\alpha(\xi)\;\bar{\Gamma}(\tilde{\theta}_n - \tilde{\theta}_{H,n}\,,\,\tilde{\theta}_n - \tilde{\theta}_{H,n})\;/\;P_n) \longrightarrow \bar{\chi}^2_{(\bar{\Gamma}^{-1},0)}\ .$$

Soit θ quelconque $\theta \in H$.

On note $P_{\theta,n}$ la mesure de probabilité de densité

$$\prod_{i=1}^{n} f_i(.,\theta,\xi)\ .$$

On ne peut montrer dans le cas général que l'hypothèse H_0 est la sous hypothèse de H asymptotiquement la plus défavorable dans le sens où le niveau asymptotique des tests, que l'on pourrait construire avec des statistiques équivalentes à $T_{\tilde{H}}$, serait maximal sous H_0. C'est-à-dire que l'inégalité

$\forall\ c > 0$

$$\lim_{n\to+\infty} P_{\theta,n}[n\,\alpha(\xi)\,\|\tilde{\theta}_n - \tilde{\theta}_{H,n}\|^2 \geq c\,] \leq$$

$$\lim_{n\to+\infty} P_n\,[n\,\alpha(\xi)\|\tilde{\theta}_n - \tilde{\theta}_{H,n}\|^2 \geq c\,]$$

n'est pas nécessairement vérifiée.

Une inégalité équivalente est vérifiée dans le modèle qu'étudient Robertson et Wegman (1978), les lois de $\bar{\chi}^2$ asymptotiques qu'obtiennent ces auteurs sont moins dépendantes de la vraie valeur du paramètre que dans notre cas. Toutefois si θ n'est pas sur la frontière de H l'inégalité est vérifiée.

Cette inégalité est aussi vérifiée, bien sûr, dans la situation où on étudie le cas normal (voir proposition I.7).
Pour ces raisonset pour des raisons de commodités on étudiera donc le comportement des statistiques de test sous H_0.

III.7.1 - Un test de H

On considère que la condition III.3 sur ξ est toujours valable et de manière analogue à ce qui a déjà été fait précédemment, on considère la statistique $T_{1,H}$ définie par

$$T_{1,H} = n\,\alpha(\hat{\xi})\,\|\hat{\theta}_n - \hat{\theta}_{H,n}\|^2$$

et on définit le test de H par

rejet de H pour $T_{1,H} \geq \ell$

soit $\phi_{1,H} = 1_{\{T_{1,H} \geq \ell\}}$

Proposition III.6 : *Sous les conditions II.3 et III.3*

$$T_{\tilde{H}} - T_{1,H} \xrightarrow{P} 0$$

La démonstration est semblable à la démonstration de la proposition III.3.
On utilise le fait que $\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) \xrightarrow{P} 0$ et $\sqrt{n}(\hat{\theta}_{H,n} - \tilde{\theta}_{H,n}) \xrightarrow{P} 0$.

III.7.2 - Test du rapport de vraisemblance de H

On définit une statistique de test par

$$T_{2,H} = 2\,[\,L_n(\cdot,\hat{\theta}_n,\hat{\xi}) - L_n(\cdot,\hat{\theta}_{H,n},\hat{\xi})]$$

et on définit le test du rapport de vraisemblance par rejet de H pour $T_{2,H} \geq \ell$.

Soit $\phi_{2,H} = 1_{\{T_{2,H} \geq \ell\}}$.

Proposition III.7 : *Sous les conditions II.3 et III.3*

$$T_{2,H} - T_{\tilde{H}} \xrightarrow{P} 0$$

pour la démonstration, voir la démonstration de la proposition III.4.

III.7.3 - Test des multiplicateurs de Lagrange.

On est conduit, pour faire les tests précédents, à obtenir une estimation de θ sous contrainte d'appartenance à un cône convexe polyhédrique, H étant toujours défini de la même manière on s'intéresse à l'estimateur du maximum de vraisemblance sous la contrainte $\theta \in H$.
Soit $L_n(\omega,\theta,\xi)$ le logarithme de vraisemblance. On cherche $\hat{\theta}_{H,n}$ tel que $L_n(\omega,\hat{\theta}_{H,n},\xi) = \max_{\theta \in H} L_n(\omega,\theta,\xi)$. Si la fonction qui à θ associe $L_n(\omega,\theta,\xi)$ est concave, c'est-à-dire si la vraisemblance est Log-concave, l'estimateur du maximum de vraisemblance sous la contrainte H sera solution du système suivant :

$$\begin{cases} L'_n(\omega,\theta,\xi) - \alpha(\xi)\ \lambda(\omega) = 0 \\ \theta \in H \end{cases}$$

où λ désigne un multiplicateur de Lagrange $\lambda(\omega) \in \mathbb{R}^m$ et toutes ses composantes sont positives c'est-à-dire que $\lambda(\omega) \in H$.

Pour plus de détails sur les multiplicateurs de Lagrange dans le cas d'appartenance à un polyèdre voir Rockafellar (1970). On peut donc se poser la question de savoir s'il est possible de construire un test de H en utilisant ce multiplicateur, car ce multiplicateur est d'autant plus "grand" que l'on est "loin" de l'hypothèse H.

Apparemment on est gêné par le fait que ces multiplicateurs n'existent que dans le cas où le logarithme de la vraisemblance est une fonction concave. Or ceci est le cas lorsque l'on affaire à une structure gaussienne et lorsque la structure n'est pas gaussienne on renonce à avoir des résultats exacts pour obtenir des résultats asymptotiques. De plus même dans le cas de la concavité du logarithme de vraisemblance la détermination simultanée de l'estimateur du maximum de vraisemblance et du multiplicateur de Lagrange peut s'avérer trop difficile, on renonce à les chercher et on considère $\hat{\theta}_{H,n}$ défini comme précedemment et on recherche une quantité $\hat{\lambda}_n$ dont la détermination sera aisée et telle que $L'_n(\omega,\hat{\theta}_{H,n},\xi) - \alpha(\xi)\hat{\lambda}_n$ converge en probabilité vers 0.
Dans le cas où le logarithme de la vraisemblance n'est pas concave on peut se demander si la quantité $\hat{\lambda}_n$, définie de la même manière que pour le cas où cette concavité est réalisée, a un sens et si elle a des propriétés asymptotiques intéressantes.

On peut définir $\tilde{\lambda}_n$ le multiplicateur de Lagrange dans la structure gaussienne tangente. On peut se demander si $\hat{\lambda}_n$ et $\tilde{\lambda}_n$ sont "proches" asymptotiquement.

Dans $\mathbb{R}^m$ et $\mathbb{R}^{m*}$ on utilise le produit scalaire induit par $\bar{\Gamma}$.

Définition : $\hat{\theta}_{H,n}$ *étant la projection de* $\hat{\theta}_n$ *sur* H *au sens de* $\bar{\Gamma}$

On définit $\hat{\lambda}_n$ *à valeur dans* $\mathbb{R}^{m*}$ *par*

$$\hat{\lambda}_n = n\bar{\Gamma}(\hat{\theta}_{H,n} - \hat{\theta}_n)$$

$\hat{\lambda}_n$ existe et est unique.
De plus comme on l'a déjà vu $\sqrt{n}$ $(\hat{\theta}_{H,n} - \hat{\theta}_n)$ a une limite en loi et $(\hat{\theta}_{H,n} - \hat{\theta}_n) \xrightarrow{P} 0$ en utilisant ces résultats et en appliquant exactement la même méthode de démonstration que Mathieu (1978 proposition II.6) on a le résultat suivant :

Proposition III.8 : *Sous les conditions II.3*

$$\frac{1}{\sqrt{n}}[L_n'(.,\hat{\theta}_{H,n},\xi) - \alpha(\xi)\hat{\lambda}_n] \xrightarrow{P} 0$$

$\tilde{\theta}_{H,n}$ désignant la projection de $\tilde{\theta}_n$ sur H au sens de $\bar{\Gamma}$.

On définit $\tilde{\lambda}_n$ par

$$\tilde{\lambda}_n = n\,\bar{\Gamma}(\tilde{\theta}_{H,n} - \tilde{\theta}_n)$$

En utilisant la même démonstration que Mathieu (1978 proposition II.7) (la seule chose qui change dans les démonstrations, c'est la signification de H dans le cas de Mathieu c'est un sous espace dans le nôtre c'est le polyèdre déjà défini). On trouve le résultat suivant :

Proposition III.9 : *Sous les conditions II.3*

$$\frac{1}{\sqrt{n}}\|\tilde{\lambda}_n - \hat{\lambda}_n\| \xrightarrow{P} 0$$

A partir de là on peut maintenant définir les tests des multiplicateurs de Lagrange.

On définit $T_{3,H}$ par

$$T_{3,H} = \frac{\alpha(\hat{\xi})}{n}\ \bar{\Gamma}^{-1}(\hat{\lambda}_n,\hat{\lambda}_n) = \frac{\alpha(\hat{\xi})}{n}\ \|\hat{\lambda}_n\|^2$$

et on définit un test de H par

rejet de H pour $T_{3,H} \geq \ell$

soit $\phi_{3,H} = 1_{\{T_{3,H} \geq \ell\}}$

On définit aussi la statistique $T_{4,H}$ par

$$T_{4,H} = \frac{1}{n} \bar{\Gamma}^{-1}[L'_n(.,\hat{\theta}_{H,n},\hat{\xi}) \, , L'_n(.,\hat{\theta}_{H,n},\hat{\xi})]$$

$$= \frac{1}{n} \|L'_n(.,\hat{\theta}_{H,n},\hat{\xi})\|^2$$

et on définit un autre test de H par

rejet de H pour $T_{4,H} \geq \ell$

soit $\phi_{4,H} \geq 1_{\{T_{4,H} \geq \ell\}}$

<u>Proposition III.10</u> : *Sous les conditions II.3 et III.3*

$$T_{3,H} - T_{\tilde{H}} \xrightarrow{P} 0$$

et

$$T_{4,H} - T_{\tilde{H}} \xrightarrow{P} 0$$

<u>Démonstration</u> : $\frac{1}{n}\|\hat{\lambda}_n\|^2 - \frac{1}{n}\|\tilde{\lambda}_n\|^2 \xrightarrow{P} 0$

$$\frac{1}{n}\|\tilde{\lambda}_n\|^2 = n\,\|\tilde{\theta}_{H,n} - \tilde{\theta}_n\|^2$$

$$\alpha(\hat{\xi}) \xrightarrow{P} \alpha(\xi)$$

d'où $\frac{\alpha(\hat{\xi})}{n} \|\hat{\lambda}_n\|^2 - n(\alpha(\xi)\,\|\tilde{\theta}_{H,n} - \tilde{\theta}_n\|^2 \xrightarrow{P} 0$

donc $T_{3,H} - T_{\tilde{H}} \xrightarrow{P} 0$

comme $\frac{1}{\sqrt{n}}[L'_n(.,\hat{\theta}_{H,n},\xi) - \alpha(\xi)\hat{\lambda}_n] \xrightarrow{P} 0$

on a aussi

$$T_{4,H} - n\,\alpha(\xi)\,\|\tilde{\theta}_{H,n} - \tilde{\theta}_n\|^2 \xrightarrow{P} 0$$

On a donc 4 tests de H, ces 4 tests sont asymptotiquement équivalents dans le sens précédemment défini. Ce résultat se montre de manière analogue à ce qui a déjà été fait pour $T_{1,H_{o,1}}$ et $T_{2,H_{o,1}}$.

CHAPITRE IV

UNE EXTENSION DES TESTS DU $\bar{\chi}^2$ AU CAS OÙ L'HYPOTHÈSE NULLE EST COMPOSÉE DANS LE G.L.I.M.

- o - o - o - o -

IV.1 - Introduction : étude dans le cas normal

(Voir Barlow et coll. 1972 p. 179-180).

On considère la structure statistique suivante :

$$(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m}, \{N_m(\mu,\Sigma),\ \mu \in \mathbb{R}^m\})^n$$

Soit g une application linéaire surjective de $\mathbb{R}^m$ dans $\mathbb{R}^q$ $q < m$

On considère les hypothèses suivantes :

H_o : $g(\mu) = 0$

H : $g(\mu) \geq 0$ c'est-à-dire toutes les composantes de $g(\mu)$ sont positives.

H_1 : $H \setminus H_o$

On veut tester H_o contre H_1

On montre facilement que l'on peut se ramener à la situation suivante :

Soit $\mu \in \mathbb{R}^m$ on écrit ${}^t\mu = (\mu_1,\ldots,\mu_m)$

$$= (\mu_1,\ldots,\mu_q,\mu_{q+1},\ldots,\mu_m)$$

on note ${}^t\mu_{[1]} = (\mu_1,\ldots,\mu_q)$

et on considère les hypothèses

H_o : ${}^t\mu_{[1]} = 0 \iff \mu_i = 0 \quad \forall i = 1,\ldots,q$

H : ${}^t\mu_{[1]} \geq 0 \iff \mu_i \geq 0 \quad \forall i = 1,\ldots,q$

$H_1 = H \setminus H_o$

et on s'intéresse au test de H_o contre H_1

il suffit pour cela de changer de matrice de variances et covariances.

On se contentera donc d'étudier cette dernière situation.

L'estimateur du maximum de vraisemblance de μ sous contrainte est bien sûr

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i .$$

La statistique du test du rapport de vraisemblance de H_o contre H_1 est

$$\bar{\chi}^2 = n \, \|Y_H - Y_{H_o}\|^2_{\Sigma^{-1}}$$

$\bar{Y}_H$ désigne la projection de $\bar{Y}$ sur H au sens de Σ^{-1}

$\bar{Y}_{H_o}$ désigne la projection de $\bar{Y}$ sur H_o au sens de Σ^{-1}

${}^t\bar{Y} = (\bar{Y}_1,\ldots,\bar{Y}_q, \bar{Y}_{q+1},\ldots,\bar{Y}_m)$

${}^t\bar{Y}_{[1]} = (\bar{Y}_1,\ldots,\bar{Y}_q)$

On pose $\Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

Σ_{11} matrice $q \times q$

Σ_{12} matrice $q \times (m-q)$

$\Sigma_{21} = {}^t\Sigma_{12}$

Λ_{11} est définie par

$$\Lambda_{11} = (\Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21})^{-1}$$

μ_1^* est la quantité qui minimise ${}^t(Y_{[1]} - \mu_1)\,\Lambda_{11}^{-1}\,(Y_{[1]} - \mu_1)$

il est montré dans Barlow et coll. (1972) que

$$\bar{X}^2 = n\,\|\mu_1^*\|^2_{\Lambda_{11}^{-1}}$$

On en déduit donc que si $\mu \in H_0$ la statistique du test du rapport de vraisemblance suit la loi $\bar{\chi}^2_{\Lambda_{11}}$ c'est-à-dire la loi de $\bar{\chi}^2$ dépendant de la matrice Λ_{11} .

On note $\Psi(\Sigma) = (\Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}^{-1})^{-1} = \Lambda_{11}$

On remarque que si $q = m$ $\Psi(\Sigma) = \Sigma$
et l'on retrouverait le test de $H_0 : \theta = 0$ contre

$H_1 = H \backslash H_0$ ou $H : \theta_i \geq 0 \quad \forall i = 1,\dots,m.$

On voit donc que si l'on se situe dans le cas des lois normales il n'y a pas de grosses difficultés supplémentaires. Toutefois pour une étude plus poussée on renvoie le lecteur à Barlow et coll (1972).

IV.2 - Etude dans le cadre du modèle linéaire généralisé

L'extension des tests d'hypothèses linéaires composées contre des hypothèses polyédriques dans le cadre du modèle linéaire généralisé pose certaines difficultés spécifiques dues essentiellement au fait que pour l'étude des tests on se place dans une structure gaussienne tangente (voir ce qui a été fait au chapitre III). Lorsque l'hypothèse est simple (en θ) on considère la structure gaussienne tangente en ce point unique et connu. Maintenant l'hypothèse nulle étant composée, on va avoir des problèmes notamment pour connaître les valeurs de rejet des tests car ces valeurs dépendront de la matrice des variances et covariances de la structure gaussienne tangente en la vraie valeur du paramètre, laquelle n'est pas connue.

On se place donc dans la structure précédemment envisagée

$$\bigotimes_{i=1}^{n} (\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}, f_i(\, ,\theta,\xi), \theta \in \Theta, \xi \in \Xi)$$

où $f_i(\cdot,\theta,\xi)$ a la forme explicitée au chapitre II voir (3) de la même manière que précédemment on s'intéresse au test de l'hypothèse

H_o : $g(\theta) = 0$ où g est linéaire surjective de $\mathbb{R}^m$ dans $\mathbb{R}^q$ contre l'hypothèse H_1

$$H_1 = H \backslash H_o$$

où H : $g(\theta) \geq 0$ c'est-à-dire toute les composantes positives.

On peut montrer aussi que l'on peut se ramener à la situation suivante

$$^t\theta = (\theta_1,\ldots,\theta_m) = {}^t(\theta_1,\ldots,\theta_q \, , \theta_{q+1} \, , \ldots, \theta_m)$$

$$^t\theta_{[1]} = (\theta_1 \, , \ldots, \theta_q)$$

H_o : $\theta_{[1]} = 0$ i.e. $\theta_i = 0$ $\forall i = 1,\ldots,q$

H : $\theta_i \geq 0$ $\forall i = 1,\ldots,q$ $\quad H_1 = H\backslash H_o$

On veut tester H_o contre H_1

En effet il existe h bijection linéaire de $\mathbb{R}^m$ dans $\mathbb{R}^m$ telle que

$$\{\theta/g(\theta) = 0\} = \{\theta/h(\theta) \text{ a ses q premières composantes nulles}\}$$

de plus dans le modèle linéaire généralisé θ intervient sous la forme $t_i\theta$ où t_i est une application linéaire connue.

On reparamétrise $t_i\theta = t_i \circ h^{-1}[h(\theta)]$

On remplace θ par $h(\theta)$

et on remplace t_i par $t_i \circ h^{-1}$

IV.3 - Problèmes d'estimation de θ sous contraintes

$\hat{\theta}_n$ désigne un des estimateurs de θ sans contrainte vérifiant les conditions II. (11).

$\tilde{\theta}_n$ désigne l'estimateur de θ dans la structure gaussienne tangente tel qu'il a été défini au chapitre II.

On considère $\overline{\theta} \in H_o$

On rappelle que l'on note

$$\Gamma(\hat{\theta}_n) = \Gamma_n \quad \text{et} \quad \Gamma(\overline{\theta}) = \overline{\Gamma}$$

Comme précédemment on ne va pas chercher explicitement l'estimation du maximum de vraisemblance sous les contraintes H_o et H. On va considérer de manière analogue à ce que l'on a déjà fait les estimateurs suivants de θ .

$\hat{\theta}_{H_{o},n}$ projection de $\hat{\theta}_n$ sur le sous-espace H_o au sens du produit scalaire défini par Γ_n.

$\hat{\theta}_{H,n}$ projection de $\hat{\theta}_n$ sur le cône polyédrique H au sens du produit scalaire défini par Γ_n.

De même on considère les estimateurs de θ dans la structure gaussienne tangente.

$\tilde{\theta}_{H_o,n}$ projection de $\tilde{\theta}_n$ sur H_o au sens de $\overline{\Gamma}$

$\tilde{\theta}_{H,n}$ projection de $\tilde{\theta}_n$ sur H au sens de $\overline{\Gamma}$

Les propriétés de $\hat{\theta}_{H_o,n}$ et $\tilde{\theta}_{H_o,n}$ ont été étudiés par Mathieu (1978, 1981) et on a notamment la proposition suivante :

<u>Proposition IV.1</u>. *Sous les conditions II.3*

$$\sqrt{n}(\hat{\theta}_{H_o,n} - \tilde{\theta}_{H_o,n}) \xrightarrow{P} 0 .$$

On a aussi le nouveau résultat suivant :

<u>Proposition IV.2</u>. *Sous les conditions II.3.*

$$\sqrt{n}(\hat{\theta}_{H,n} - \tilde{\theta}_{H,n}) \xrightarrow{P} 0$$

<u>Démonstration</u> :

Pour les besoins de la démonstration, on va introduire un nouvel "estimateur" de θ .

$\theta^*_{H,n}$ est la projection de θ_n sur H au sens du p.s. $\overline{\Gamma}$, en fait plus qu'un estimateur $\theta^*_{H,n}$ est une variable de travail.

On va montrer dans un premier temps que $\sqrt{n}(\hat{\theta}_{H,n}-\theta^*_{H,n}) \xrightarrow{P} 0$

H est un convexe fermé, $\hat{\theta}_{H,n}$ étant la projection de $\hat{\theta}_n$ sur H au sens de Γ_n , $\hat{\theta}_{H,n}$ est caractérisé par l'inégalité suivante :

$\forall h \in H \quad {}^t(\hat\theta_n-\hat\theta_{H,n}) \quad \Gamma_n(h-\hat\theta_{H,n}) \leq 0$ $\qquad (C_1)$

de même $\theta^*_{H,n}$ étant la projection de $\hat\theta_n$ sur le convexe H au sens de $\overline{\Gamma}$ $\theta^*_{H,n}$ est caractérisé par l'inégalité suivante :

$\forall h \in H \quad {}^t(\hat\theta_n-\theta^*_{H,n}) \,\overline{\Gamma}(h-\theta^*_{H,n}) \leq 0$ $\qquad (C_2)$

On calcule la quantité suivante :

$$\| \hat\theta_{H,n} - \theta^*_{H,n} \|^2_{\Gamma_n} = {}^t(\hat\theta_{H,n}-\theta^*_{H,n}) \,\Gamma_n(\hat\theta_{H,n}-\theta^*_{H,n}) =$$

$$= {}^t(\hat\theta_{H,n}-\hat\theta_n) \,\Gamma_n(\hat\theta_{H,n}-\theta^*_{H,n}) + {}^t(\hat\theta_n-\theta^*_{H,n}) \,\Gamma_n(\hat\theta_{H,n}-\theta^*_{H,n})$$

Comme $\theta^*_{H,n} \in H$ d'après C_1 on a

${}^t(\hat\theta_{H,n}-\hat\theta_n) \,\Gamma_n(\hat\theta_{H,n}-\theta^*_{H,n}) \leq 0$ d'où l'on déduit l'inégalité suivante :

$$\|\hat\theta_{H,n} - \theta^*_{H,n} \|^2_{\Gamma_n} \leq {}^t(\hat\theta_n-\theta^*_{H,n}) \,\Gamma_n(\hat\theta_{H,n}-\theta^*_{H,n}) =$$

$${}^t(\hat\theta_n-\theta^*_{H,n})(\Gamma_n-\overline{\Gamma})(\hat\theta_{H,n}-\theta^*_{H,n}) + {}^t(\hat\theta_n-\theta^*_{H,n})\overline{\Gamma}(\hat\theta_{H,n}-\theta^*_{H,n})$$

Comme $\hat\theta_{H,n} \in H$ on a d'après C_2

$${}^t(\hat\theta_n-\theta^*_{H,n}) \,\overline{\Gamma}(\hat\theta_{H,n}-\theta^*_{H,n}) \leq 0$$

d'où l'on déduit

$$\| \hat\theta_{H,n}-\theta^*_{H,n}\|^2_{\Gamma_n} \leq {}^t(\hat\theta_n-\theta^*_{H,n})(\Gamma_n-\overline{\Gamma})(\hat\theta_{H,n}-\theta^*_{H,n})$$

On note $\| \cdot \|_e$ la norme euclidienne ordinaire sur $\mathbb{R}^m$ on munit l'espace des matrices d'une norme quelconque.

On a l'inégalité suivante :

$$\| \hat\theta_{H,n}-\theta^*_{H,n} \|^2_{\Gamma_n} \leq \text{Cte} \,\|\Gamma_n-\overline{\Gamma}\| \;\|\hat\theta_n-\theta^*_{H,n}\|_e \,\|\hat\theta_{H,n}-\theta^*_{H,n}\|_e$$

la constante Cte ne dépend que du choix de la norme sur l'espace des matrices.

On a aussi la relation suivante

$$\| \hat\theta_{H,n}-\theta^*_{H,n}\|^2_e \leq \frac{\| \hat\theta_{H,n}-\theta^*_{H,n} \|^2_{\Gamma_n}}{\lambda_n^{(1)}}.$$

où $\lambda_n^{(1)}$ désigne la plus petite valeur propre de la matrice Γ_n.

Les 2 inégalités précédentes impliquent :

$$\| \hat{\theta}_{H,n}-\theta^*_{H,n}\|^2_{\Gamma_n} \leq \frac{Cte}{\sqrt{\lambda_n^{(1)}}} \| \Gamma_n-\overline{\Gamma} \| \|\hat{\theta}_n-\theta^*_{H,n}\|_e$$

$\lambda_n^{(1)}$ existe et est strictement positif car Γ_n est supposée définie positive.

$\sqrt{n}(\hat{\theta}_n-\overline{\theta})$ a une limite en loi

$\theta^*_{H,n}$ projection de $\hat{\theta}_n$ sur H au sens de $\overline{\Gamma}$

comme $\overline{\theta} \in H_o$

$(\theta^*_{H,n}-\overline{\theta})$ est la projection de $(\hat{\theta}_n-\overline{\theta})$ sur H au sens de $\overline{\Gamma}$.

On note H^o le cône polaire de H pour le p.s. $\overline{\Gamma}$

$(\hat{\theta}_n-\theta^*_{H,n})$ est la projection de $(\hat{\theta}_n-\overline{\theta})$ sur H^o au sens de $\overline{\Gamma}$ d'où

$\sqrt{n}(\hat{\theta}_n-\theta^*_{H,n})$ a une limite en loi

$$\hat{\theta}_n \xrightarrow{P} \overline{\theta}$$

$$\Gamma_n \xrightarrow{P} \overline{\Gamma} \implies \|\Gamma_n-\overline{\Gamma}\| \xrightarrow{P} 0$$

On en déduit que

$$\frac{Cte}{\sqrt{\lambda_n^{(1)}}} \|\Gamma_n-\overline{\Gamma}\| \, \|\sqrt{n}(\hat{\theta}_n-\theta^*_{H,n})\| \xrightarrow{P} 0$$

d'où $\sqrt{n}(\hat{\theta}_{H,n}-\theta^*_{H,n}) \xrightarrow{P} 0$

de plus on a le résultat suivant

$$\sqrt{n}(\hat{\theta}_n-\tilde{\theta}_n) \xrightarrow{P} 0$$

La projection sur le convexe H au sens $\overline{\Gamma}$ étant contractante

$$\sqrt{n}(\theta^*_{H,n}-\tilde{\theta}_{H,n}) \xrightarrow{P} 0$$

il s'ensuit que :

$$\sqrt{n}(\hat{\theta}_{H,n}-\tilde{\theta}_{H,n}) \xrightarrow{P} 0$$

<u>Corollaire</u> : *Sous les conditions II.3*

$$(\hat{\theta}_{H,n}-\hat{\theta}_n) \xrightarrow{P} 0$$

En effet $[\sqrt{n}(\hat{\theta}_{H,n}-\hat{\theta}_n)-\sqrt{n}(\theta^*_{H,n}-\hat{\theta}_n)] \xrightarrow{P} 0$

Comme $\sqrt{n}(\theta^*_{H,n}-\hat{\theta}_n)$ converge en loi

$\sqrt{n}(\hat{\theta}_{H,n}-\hat{\theta}_n)$ converge en loi

et donc $(\hat{\theta}_{H,n}-\hat{\theta}_n) \xrightarrow{P} 0$

On considère que les conditions III.3 sur ξ sont valables et on définit la statistique suivante :

$$T_{H_{o,1}} = n\alpha(\xi) \, \|\tilde{\theta}_{H,n}-\tilde{\theta}_{H_o,n}\|^2$$

On va d'abord établir la distribution asymptotique de $T_{\tilde{H}o,1}$ ensuite on construira des tests de H_o contre H_1 à partir de statistiques asymptotiquement équivalentes à $T_{\tilde{H}_{o,1}}$

IV.4. Distribution de $T_{\tilde{H}_{o,1}}$ sous P_n

On sait que $\mathcal{L}(\sqrt{n}(\tilde{\theta}_n-\overline{\theta})/P_n) \to N(0,\frac{1}{\alpha(\xi)} \; \overline{\Gamma}^{-1})$

$$\tilde{\theta}_{H,n}-\tilde{\theta}_{H_o,n} = (\tilde{\theta}_{H,n}-\overline{\theta}) - (\tilde{\theta}_{H_o,n}-\overline{\theta})$$

$\tilde{\theta}_{H,n}-\overline{\theta}$ est la projection de $\tilde{\theta}_n-\overline{\theta}$ sur H_o au sens de $\overline{\Gamma}$ car $\overline{\theta}$ H_o.

de même

$\tilde{\theta}_{H,n}-\overline{\theta}$ est la projection de $\tilde{\theta}_n-\overline{\theta}$ sur H au sens de $\overline{\Gamma}$.

Proposition IV.3. *$X_n \longrightarrow N_m(0,\Sigma)$*

$X_{H,n}$ projection de X_n sur H au sens de Σ^{-1}

$X_{H_o,n}$ projection de X_n sur H_o au sens de Σ^{-1}

$$\|X_{H,n}-X_{H_o,n}\|^2_{\Sigma^{-1}} \longrightarrow \bar{\chi}^2_{\Psi(\Sigma)}$$

La démonstration se fait comme celle de la proposition III.2 On utilise la définition de $\bar{\chi}^2_{\Psi(\Sigma)}$ du IV.1 .

On déduit de cette proposition le résultat suivant

$$\mathcal{L}(T_{\tilde{H}_{o,1}}/P_n) \longrightarrow \bar{\chi}^2_{\Psi(\overline{\Gamma}^{-1})}.$$

Remarques : On pourrait définir, comme dans le cas du test d'une hypothèse simple, ce qu'est une loi de $\bar{\chi}^2$ décentrée obtenue dans le cas du test d'une hypothèse nulle composée.

On pourrait à partir de là trouver la distribution asymptotique de $T_{\tilde{H}_{o,1}}$ sous P'_n et à partir de là étudier les "puissances" asymptotiques des tests de H_o contre H_1. Ce travail soulève de gros problèmes de calcul.

On constate dans l'expression de la loi asymptotique de $T_{\tilde{H}_{o,1}}$ que, sous hypothèse nulle, cette loi dépend de la matrice $\bar{\Gamma}$ qui elle dépend de la vraie valeur du paramètre $\bar{\theta}$ donc si $\ell(\alpha,\bar{\theta})$ est défini par

$$P(\bar{X}^2_{\psi(\bar{\Gamma}^{-1})} \geq \ell_{\alpha,\bar{\theta}}) = \alpha$$

$\ell_{\alpha,\bar{\theta}}$ est en général inconnu dépendant de $\bar{\theta}$.

On définit une statistique de test de la manière suivante

$$\begin{aligned} T_{1,H_{o,1}} &= n\alpha(\hat{\xi})\, \Gamma_n(\hat{\theta}_{H,n}-\hat{\theta}_{H_o,n},\hat{\theta}_{H,n}-\hat{\theta}_{H_o,n}) \\ &= n\alpha(\hat{\xi})\, \|\hat{\theta}_{H,n}-\hat{\theta}_{H_o,n}\|^2_n \end{aligned}$$

Proposition IV.4. *Sous les conditions II.3. et III.3.*

$$T_{1,H_{o,1}} - T_{\tilde{H}_{o,1}} \xrightarrow{P} 0$$

Démonstration :

$$\alpha(\hat{\xi}) \xrightarrow{P} \alpha(\xi)$$

$$\Gamma_n \xrightarrow{P} \bar{\Gamma}$$

$$\sqrt{n}(\hat{\theta}_{H,n}-\tilde{\theta}_{H,n}) \xrightarrow{P} 0$$

$$\sqrt{n}[(\hat{\theta}_{H,n}-\hat{\theta}_{H_o,n})-(\tilde{\theta}_{H,n}-\tilde{\theta}_{H_o,n})] \xrightarrow{P} 0$$

donc

$$n\alpha(\hat{\xi})\Gamma_n(\hat{\theta}_{H,n}-\hat{\theta}_{H_o,n},\hat{\theta}_{H,n}-\hat{\theta}_{H_o,n})-n\alpha(\xi)\, \bar{\Gamma}(\tilde{\theta}_{H,n}-\tilde{\theta}_{H_o,n},\tilde{\theta}_{H,n}-\tilde{\theta}_{H_o,n}) \xrightarrow{P} 0$$

d'où le résultat.

On ne peut pas pour les raisons expliquées plus haut construire un test de la manière suivante :

rejet de H_o pour $T_{1,H_{o,1}} \geq \ell_{\alpha,\bar\theta}$

soit $\phi_{1,H_{o,1}} = 1\!\!1\{T_{1,H_{o,1}} \geq \ell_{\alpha,\bar\theta}\}$

On définit aussi une statistique du test de rapport de vraisemblance par

$$T_{2,H_{o,1}} = 2[L_n(.,\hat\theta_{H,n},\hat\xi) - L_n(.,\hat\theta_{H_{o,n}},\hat\xi)]$$

Proposition IV.5. : *Sous les conditions II.3 et III.3*

$$T_{2,H_{o,1}} - T_{\tilde H_{o,1}} \xrightarrow{P} 0$$

La démonstration de ce résultat est similaire à celle de la proposition III.4.

On ne peut pas, ici non plus, construire un test par rejet de H_o pour $T_{2,H_{o,1}} \geq \ell_{\alpha,\bar\theta}$

L'idée qui vient pour lever cette difficulté est d'utiliser un estimateur de $\ell_{\alpha,\bar\theta}$. Comment construire un estimateur de $\ell_{\alpha,\bar\theta}$ convenable ? Comme l'on dispose d'un estimateur de $\bar\theta$, $\hat\theta_n$ on peut se demander si $\ell_{\alpha,\hat\theta_n}$ défini de la manière suivante :

$P(\bar X^2_{\psi(\Gamma_n^{-1})} \geq \ell_{\alpha,\hat\theta_n}) = \alpha$ n'est pas un "bon" estimateur de $\ell_{\alpha,\bar\theta}$.

Pour étudier la qualité de $\ell_{\alpha,\hat\theta_n}$ on a besoin de résultats préliminaires.

IV.5. Continuité de la probabilité orthante.

Soit l'application qui à Σ matrice symétrique définie positive associe $P(\Sigma)$ la probabilité orthante pour la loi normale $N(0,\Sigma)$ telle qu'elle a déjà été définie.

On rappelle

$$P(\Sigma) = (2\pi)^{-k/2}|\Sigma|^{-1/2} \int_0^{+\infty}\int_0^{+\infty} \cdots \int_0^{+\infty} \exp - \frac{1}{2}\, {}^t x\, \Sigma^{-1} x\, dx$$

Pour le calcul numérique de $P(\Sigma)$ on renvoie aux papiers de A. Kudô (1963), Bohrer (1975), Milton (1972), J.E. Dutt et T.K. Linn (1975).

Proposition IV.6. : *L'application $\Sigma \longrightarrow P(\Sigma)$ de l'ensemble des matrices symétriques définies positives $k \times k$ dans $\mathbb{R}$ est continue.*

Démonstration : On note $f(x,\Sigma) = (2\pi)^{-k/2}\ |\Sigma|^{-1/2}\ \exp - \frac{1}{2}\ {}^t x\ \Sigma^{-1}\ x$

$P(\Sigma) = \int_H f(x,\Sigma)\ dx$

On a : $\int_{\mathbb{R}^k} f(x,\Sigma)\ dx = 1$

On considère une suite $(\Sigma_n)_{n\in\mathbb{N}}$ de matrices symétriques définies positives telles que $\Sigma_n \to \Sigma$

$$\int_{\mathbb{R}_k} f(x,\Sigma_n)dx = 1$$

l'application $\Sigma \to f(x,\Sigma)$ est continue

$f(x,\Sigma_n)$ converge ponctuellement vers $f(x,\Sigma)$.

On peut appliquer le lemme de Scheffé voir Lehmann (1959)

$f(.,\Sigma_n)$ converge en moyenne vers $f(.,\Sigma)$

en appliquant un autre résultat de Lehmann (1959), on trouve que

$$\int_H f(x,\Sigma_n)\ dx \longrightarrow \int_H f(x,\Sigma)\ dx$$

c'est-à-dire $\quad P(\Sigma_n) \longrightarrow P(\Sigma)$

On en déduit la continuité de P.

IV.6. Résultats complémentaires

On définit $\gamma_j(\Sigma)$ de la façon suivante :

$$P(\bar{X}^2_\Sigma \geq c) = \sum_{j=0}^{k} \gamma_j(\Sigma)\ \ P(\chi^2_j \geq c)$$

Proposition IV.7 : *L'application de l'ensemble des matrices symétriques définies positives dans $\mathbb{R}^{k+1}$*

$\Sigma \to \Sigma(\gamma_0(\Sigma),\gamma_1(\Sigma),\ldots,\gamma_k(\Sigma))$ est continue.

Cela découle immédiatement de la définition de $\bar{\gamma}_j(\Sigma)$ de la proposition IV.6. et de l'expression

$$P(\bar{\chi}^2_\Sigma \geq c) = \sum_{\emptyset\subset M\subset K} P(\chi^2_{n(M)} \geq c)\ P((\Sigma_{M'})^{-1})\ P(\Sigma_{M,M'}).$$

Soit $\gamma = (\gamma_0,\gamma_1,\ldots,\gamma_k) \in \mathbb{R}^{k+1}$

tels que $0 \leq \gamma_i \leq 1 \quad \forall i = 0,1,\ldots,k$

$$\sum_{i=0}^{k} \gamma_i = 1$$

$$\gamma_k > 0$$

On note $\bar{\chi}^2_\gamma$ la loi de $\bar{\chi}^2$ dépendant du paramètre γ de la manière suivante $P(\bar{\chi}^2_\gamma \geq c) = \sum_{j=0}^{k} \gamma_j P(\chi^2_j \geq c)$

et on définit $c_{\alpha,\gamma}$ par

$$P(\bar{\chi}^2_\gamma \geq c_{\alpha,\gamma}) = \alpha$$

<u>Proposition IV.8.</u> : *Si α est donné l'application de $\mathbb{R}^{k+1}$ dans $\mathbb{R}$*

$$\gamma \to c_{\alpha,\gamma} \text{ est continue}$$

<u>Démonstration</u> :

$$P(\bar{\chi}^2_\gamma \geq c) = \sum_{j=0}^{k} \gamma_j P(\chi^2_j \geq c) = G(\gamma,c)$$

. α est supposé différent de 1 car les tests de niveau 1 n'ont absolument aucun intérêt. $c_{\alpha,\gamma}$ sera toujours différent de 0.
G(γ,c) est une fonction continûment différentiable de γ et de c.
$\gamma \in \mathbb{R}^{k+1}$, $c \in]0,+\infty[$.

$$\frac{\partial}{\partial c} G(\gamma,c) \leq 0$$

en appliquant le théorème des fonctions implicites voir Carton (1967) on trouve que $c_{\alpha,\gamma}$ est une fonction continue de γ.

<u>Proposition IV.9.</u> : *Sous les conditions II.3.*

$$\ell_{\alpha,\hat{\theta}_n} \xrightarrow{P} \ell_{\alpha,\bar{\theta}}$$

<u>Démonstration</u> : Il découle des propositions IV.7. et IV.8. que l'application de Θ dans $\mathbb{R}$ qui à θ associe $\ell_{\alpha,\theta}$ est continue en tant que composée d'applications continues. Comme $\hat{\theta}_n \xrightarrow{P} \bar{\theta}$ on déduit le résultat.

IV.7 - Un test de H_o contre H_1

On construit un test de H_o contre H_1 de la manière suivante : rejet de H_o pour $T_{1,H_{o,1}} \geq \ell_{\alpha,\hat{\theta}_n}$.

Soit $\phi_{1,H_{o,1}}$: $\mathbb{1}_{\{T_{1,H_{o,1}} \geq \ell_{\alpha,\hat{\theta}_n}\}}$

Le test ainsi construit est de niveau asymptotique constant égal à α ; en effet :

$$\left.\begin{array}{l} \mathcal{L}(T_{\tilde{H}_{o,1}}/P_n) \longrightarrow \chi^2_\Psi(\bar{\Gamma}^{-1}) \\ \ell_{\alpha,\hat{\theta}_n} \xrightarrow{P} \ell_{\alpha,\bar{\theta}} \end{array}\right\} \Longrightarrow$$

$$\mathcal{L}(T_{\tilde{H}_{o,1}} - \ell_{\alpha,\hat{\theta}_n}/P_n) \longrightarrow \bar{\chi}^2_\Psi(\bar{\Gamma}^{-1}) - \ell_{\alpha,\bar{\theta}}$$

$$\lim_{n\to+\infty} P_n(T_{\tilde{H}_{o,1}} \geq \ell_{\alpha,\hat{\theta}_n}) = \lim_{n\to+\infty} P_n(T_{\tilde{H}_{o,1}} - \ell_{\alpha,\hat{\theta}_n} \geq 0) =$$

$$P(\bar{\chi}^2_\Psi(\bar{\Gamma}^{-1}) - \ell_{\alpha,\bar{\theta}} \geq 0) = P(\bar{\chi}^2_\Psi(\Gamma^{-1}) \geq \ell_{\alpha,\bar{\theta}}) = \alpha$$

D'après la proposition IV.4. $T_{1,H_{o,1}}$ est asymptotiquement distribué comme la variable $T_{\tilde{H}_{o,1}}$ et donc $\phi_{1,H_{o,1}}$ est bien de niveau asymptotique α.

IV.8. - Test du rapport de vraisemblance

On définit ce test par

$$\phi_{2,H_{o,1}} = \mathbb{1}_{\{T_{2,H_{o,1}} \geq \ell_{\alpha,\hat{\theta}_n}\}}$$

Remarques :

. Les 2 tests $\phi_{1,H_{o,1}}$ et $\phi_{2,H_{o,1}}$ sont asymptotiquement équivalents

. Il existe des cas où $\ell_{\alpha,\theta}$ peut être connu, par exemple si $\forall\theta \in H_o \quad \ell_{\alpha,\theta} = \ell_\alpha$.

$\hat{\theta}_n \xrightarrow{P} \bar{\theta}$ veut dire aussi $\hat{\theta}_n \xrightarrow{P'_n} \bar{\theta}$

on a aussi donc $\ell_{\alpha,\hat{\theta}_n} \xrightarrow{P'_n} \ell_{\alpha,\bar{\theta}}$

$$\lim_{n\to+\infty} P_n'(T_{\tilde{H}_{o,1}} \geq \ell_{\alpha,\hat{\theta}_n}) = \lim_{n\to+\infty} P_n'(T_{\tilde{H}_{o,1}} \geq \ell_{\alpha,\overline{\theta}})$$

Cela veut dire que les tests construits avec $\ell_{\alpha,\hat{\theta}_n}$ sont asymptotiquement aussi "puissants" que les tests que l'on pourrait construire avec $\ell_{\alpha,\overline{\theta}}$ si cette quantité était connue.

BIBLIOGRAPHIE

[1] Barlow, R.E.- Bartholomew, D.J.- Bremner, J.M. and Brunk, H.D. (1972)- *Statistical inference under order restrictions*. London, Wiley.

[2] Bartholomew, D.J. (1959 a)- A *test of homogeneity for ordered alternatives*. Biometrika 46, 35-48.

[3] Bartholomew, D.J. (1959 b)- A *test of homogeneity for ordered alternatives II*. Biometrika 46, 328-335.

[4] Bremner, J.M. (1982)- *An algorithm for Nonnegative least squares and projection onto cones*. Compstat 82 , 155-160, Physica-Verlag, Wien.

[5] Billingsley, P. (1968)- *Convergence of probability measures*. Wiley, New-York.

[6] Bohrer, R. (1975)- *Algorithm AS 90 one sided multi-variable inference*. Appl. Statist. 27, 100-104.

[7] Cartan, H. (1967)- *Calcul differentiel dans les espaces de Banach*. Hermann, Paris.

[8] Dutt, J.E. and Lin, T.K. (1975)- A *short table for computing normal orthant probability of dimension 4 and 5*. J. Statist. Comput. Simul. 4, 95-120.

[9] Kudô, A. (1963)- A *multivariate analogue of the one sided test*. Biometrika. 50, 403-418.

[10] Kudô, A. and Fujisawa, H. (1964)- *A bivariate normal test with two sided alternative* . Math. Fac. Sci. Kyushu Univ. (A) 18, 104-108.

[11] Kudô, A. and Choï J.R. (1975)- A *generalized multivariate analogue of the one sided test*. Math. Fac. Sci. Kyushu Univ. (A)29, 303-328.

[12] Lehman E.L. (1966)- *Testing statistical hypotheses*. Wiley, New-York.

[13] Mathieu, J.R. (1978)- *Contribution à l'étude de la séparabilité des hypothèses au sens du test du* χ^2 *dans la théorie asymptotique*. Thèse, Université Paul Sabatier, Toulouse.

[14] Mathieu, J.R. (1981)- *Test of* χ^2 *in the generalized linear model*. Math. Operationsforsch. Statist., Ser. Statistics 12, 529-550.

[15] Milton, R.C. (1972)- *Computer evaluation of the multivariate normal integral*. Technometrics 14, 881-889.

[16] Nelder, J.A. and Wederburn, R.W.M. (1972)- *Generalized linear models*. J. Roy. Statist. Soc. B.35, 370-384.

[17] Nüesch, P.E. (1966)- *On the problem of testing location in multivariate populations for restricted alternatives*. Ann. Math. Statist. 37, 113-119.

[18] Philippou A.N. and Roussas, G.G. (1973)- *Asymptotic distribution of the likelihood function in the independent not identically distributed case*. Ann. Stat. 1, 454-471.

[19] Philippou, A.N. and Roussas, G.G. (1975)- *Asymptotic normality of the maximum likelihood estimate in the independent not identically distributed case*. Ann. of the Inst. of Statistic. math. 27, 45-55.

[20] Robertson, T. and Wegman, E.J. (1978)- *Likelihood ratio tests for order restrictions in exponential families*. Annals of statistics 1978, 485-505.

[21] Rockfellar, R.T. (1970)- *Convex Analysis*. Princeton University Press.

[22] Roussas, G.G. (1972)- *Contiguity of probability measures, some applications in statistics*. cambridge University Press.