ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Mathématiques

GUY FOURT

Au-delà de l'analyse de la variance ?

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 65, série *Mathématiques*, n° 15 (1977), p. 76-95

<http://www.numdam.org/item?id=ASCFM_1977__65_15_76_0>

AU-DELA DE L'ANALYSE DE LA VARIANCE ?

Guy FOURT

Université de Clermont II

-:-:-:-:-:-:-:-:-:-:-

Sommaire

Nous donnons ici (IV - VIII) une application de la théorie des cônes homogènes au problème fondamental de l'analyse de la variance multi-dimensionnelle ; nous en recherchons des solutions "optimales" (de façon plus précise les plus puissantes ou à défaut localement les plus puissantes) invariantes sous le groupe linéaire. Si les résultats de IV ne sont pas nouveaux (on les trouve déjà en filigranne dans [6]) ils ne sont malheureusement pas suffisamment connus. L'efficacité de la théorie des cônes homogènes pour la solution du problème de Hotelling nous a incité à aborder des problèmes plus généraux. On retrouvera (théorèmes 5 et 8) les bases de l'analyse de la variance unidimensionnelle hors de toute heuristique (pour leur développement on consultera [2] ou [7]) ainsi que le test de Lawley-Hotelling comme forme limite (théorème 6). Signalons enfin qu'on trouve dans [7] les mêmes tests présentés comme UMP invariants, alors que le présent article ne prouve que la propriété UMP locale. Cela paraît plus attirant mais provient de ce que le groupe de transformations considéré est plus riche, et que la famille des statistiques à l'intérieur de laquelle a été faite l'étude est beaucoup plus restreinte que dans notre présentation.

I- Introduction

De nombreux problèmes de tests sont posés de la manière suivante. Soient (X_i) $i=1,2,\ldots\rho$ un échantillon de variables aléatoires à valeurs dans R^k, laplaciennes, d'espérance m et de covariance Λ. Cette dernière matrice est inconnue et on recherche des informations sur m. En fait on peut supposer Λ inversible ; sinon les X_i appartiennent presque sûrement à un hyperplan et il convient de se placer immédiatement dans cet hyperplan pour faire l'étude. Remarquer ensuite que l'absence d'information sur Λ provient du fait suivant : des mesures ont été effectuées sur une population, sans estimation possible a priori des corrélations entre les grandeurs mesurées, celles-ci étant chois simplement pour des raisons de commodité d'expérience, et les données du problème sont moins les X_i que les $A\,(X_i)$ où A est une transformation linéaire inversible. Aussi il est raisonnable de chercher des solutions en terme de statistiques invariantes sous le groupe linéaire.

II- Pourquoi les cônes homogènes ?

Considérons les statistiques

$$\overline{X} = \frac{1}{\rho} \sum_{i=1}^{\rho} X_i$$

$$S = \sum_{i=1}^{\rho} X_i\ {}^tX_i - \rho\, \overline{X}\ {}^t\overline{X}$$

où les X_i sont des matrices colonnes d'ordre k et où ${}^tA$ désigne la transposée de la matrice A. S est une matrice symétrique. Elles forment deux statistique exhaustives. Pour des raisons d'homogénéité, nous utiliserons dans la suite $X = \sqrt{\frac{\rho}{2}}\ \overline{X}$ de préférence à $\overline{X}$.

Pour $n = \frac{k\,(k+1)}{2}$, soit E l'espace vectoriel de dimension n de matrice (k, k) symétriques, et $\mathcal{M}_k^+$ le cône des matrices définies positives, $\overline{\mathcal{M}_k^+}$ désignant le cône des matrices semi-définies positives. Soit A une matrice (k, k) inversible. On lui associe la transformation g_A définie sur E par

$$g_A(s) = A\, s^t A \qquad\qquad \forall\, s \in E$$

Cette transformation, opérant sur E, est linéaire. En fonction de A, elle n'a pas de propriétés de linéarité, mais il est immédiat que $g_A \circ g_B = g_{AB}$ et que g_A applique $\mathcal{M}_k^+$ (resp. $\overline{\mathcal{M}_k^+}$) dans lui-même. Il est également simple de vérifier que pour tout s dans $\mathcal{M}_k^+$ il existe au moins une matrice A (k,k) inversible telle que

$$s = A\,{}^tA = g_A\,(I)$$

et donc que pour tout s et s' de $\mathcal{M}_k^+$ il existe une matrice $A_{s,s'}$ telle que $s' = g_{A_{s,s'}}(s)$. Désignant par $\mathcal{G}_o$ le groupe des opérateurs g_A , $\mathcal{M}_k^+$ forme un espace homogène du groupe $\mathcal{G}_o$ et l'étude des statistiques invariantes se ramènera à l'étude des actions des éléments de la forme (A, g_A) du groupe produit $\mathcal{L}(R^k) \otimes \mathcal{G}_o$ sur l'espace homogène $R^k \otimes \mathcal{M}_k^+$ auquel appartient le couple de statistiques (X, S).

Notons que la connaissance de la donnée $X\,{}^tX$ est équivalente à celle de X (à une symétrie par rapport à l'origine près, qui ne perturbe pas S). Aussi ce problème se ramènera à l'étude des actions des éléments diagonaux de $\mathcal{G}_o \otimes \mathcal{G}_o$ sur l'espace homogène $\overline{\mathcal{M}_k^+} \otimes \mathcal{M}_k^+$.

De manière générale on dit qu'un cône C de R^n est homogène si :

- C est un ouvert ne contenant pas de droites
- il existe un groupe $\mathcal{G}_o$ de transformations linéaires sur R^n

conservant C (pour tout s et s' de C, l'équation $s' = g(s)$ a une solution dans g_o).

On montre qu'il existe alors au moins un sous groupe $\mathcal{G}$ de $\mathcal{G}_o$ opérant de façon simplement transitive sur C (c'est-à-dire que pour s et s' dans C , l'équation $s' = g(s)$ a une solution <u>unique</u> dans $\mathcal{G}$). De plus, les opérateurs de g sont représentables dans une base convenablement choisie par des matrices (n,n) triangulaires supérieures. Pour le choix d'une unité e de C on définit alors une loi de groupe $\perp$ en posant pour $s = g(e)$, $g \in G$

$$s \perp s' = g\,(s')$$

Noter que la définition de la loi $\perp$ repose sur les choix successifs de $\mathcal{G}$ et de e ... En ce qui concerne le cône homogène $\mathcal{M}_k^+$, nous utiliserons celle qui s'exprime le plus simplement, obtenue avec $\mathcal{G}$, groupe des transformations g_T, où T est une matrice (k,k) triangulaire supérieure et où e est la matrice identité sur R^k. On a donc, pour $s = T\,{}^tT$, (T triangulaire k, k)

$$s \perp s' = T\, s'\, {}^tT$$

Pardonnez-moi si la notation e se superpose avec celle de e, base des logarithmes népériens. Leur coexistence, même dans la même formule ne pose pas de soucis tant ces deux êtres sont distincts. Mon expérience m'a habitué à reserver la notation exp aux opérateurs, compte tenu de l'importance de la structure de $\mathcal{G}$ en tant que groupe de Lie.

III- Anatomie de $\mathcal{M}_k^+$

Le lecteur consultera utilement [4] où il trouvera une présentation sommaire des cônes homogènes ; pour une étude détaillée, nous le renvoyons à [5] ou [8] . Nous donnons ici un glossaire qui, nous l'espérons, lui permettra de suivre les calculs des chapitres IV à VIII.

Loi Δ : pour tout vecteur a de l'espace E , l'équation L(e) = a fournit une solution unique L_a dans l'algèbre de Lie, dérivée de G . De façon similaire, à la définition de $\perp$, on pose

$$a \,\Delta\, b = L_a\,(b) \qquad\qquad (a, b \in E)$$

Δ est une loi d'algèbre (non associative) assez sophistiquée. Mentionnons cependant les propriétés que nous utiliserons par la suite. Si M_i (i=1,2) sont des vecteurs de R^{k-1}, pour les matrices $\begin{pmatrix} 0 & M_i \\ {}^tM_i & 0 \end{pmatrix}$, notées encore M_i par abus d'écriture, Δ vérifie la relation simple :

$$M_1 \, \Delta \, M_2 = \begin{pmatrix} M_1 \, {}^tM_2 + M_2 \, {}^tM_1 & 0 \\ 0 & 0 \end{pmatrix}$$

Le lecteur en déduira sans peine que ($M_1 = M_2 = M$), $M \, \Delta \, M$ est dans $\overline{\mathcal{M}^+_{k-1}}$ ainsi que le principe de décomposition : la matrice

$$U = \begin{pmatrix} s & N \\ {}^tN & z \end{pmatrix} \qquad (N \in R^{k-1} \quad , \quad z \in R)$$

est dans $\mathcal{M}^+_k$ si et seulement si $z > 0$ et $\begin{pmatrix} s & 0 \\ 0 & 0 \end{pmatrix} - \dfrac{N \, \Delta \, N}{2z}$ est dans $\mathcal{M}^+_{k-1}$. De plus $\dfrac{N \, \Delta \, N}{2z} + N + \begin{pmatrix} 0 & 0 \\ 0 & z \end{pmatrix}$ est dans $\overline{\mathcal{M}^+_k}$.

On posera également, pour $s = T \, {}^tT$, T matrice triangulaire supérieure et $M \in R^k$, $s \perp (M) = T\,(M)$.

<u>Produit scalaire et dualité</u>. On munira R^n d'un produit scalaire compatible avec Δ au sens suivant : $< e \, , \, a \, \Delta \, b > \; = \; < a, \, b >$. Nous utiliserons ici le produit scalaire usuel de la trace qui est compatible. Le cone dual $(\mathcal{M}^+_k)^*$ de $\mathcal{M}^+_k$ n'est autre que $\mathcal{M}^+_k$, mais il faut reordonner la base dans laquelle ont été exprimées les données pour que le groupe des transformations duales g^*_T (pour $< \, , \, >$) puisse être lu comme un groupe triangulaire supérieur. Si $s = T \, {}^tT = g_T\,(e)$ (T matrice triangulaire supérieure), $s^* = {}^tT \, T = g^*_T\,(e)$ définit le point dual de s.

<u>Fonctions puissance ρ</u>

Ce sont les homomorphismes de $\mathcal{M}^+_k$ muni de $\perp$ dans R^+. Notons par $\Delta_i(s)$ le déterminant de la matrice obtenue en rayant les $i-1$ premières lignes et colonnes de s et posons $\chi_i(s) = \dfrac{\Delta_i(s)}{\Delta_{i+1}(s)}$, $\Delta_{k+1} = 1$. Tout homomorphisme s'écrit de façon canonique $\prod\limits_{i=1}^{k} (\chi_i(s))^{\rho_i}$ que nous abrégerons en $s^{(\rho_1 \ldots \rho_k)}$

Noter que $s^{*(\rho_1 \, \ldots \, \rho_k)} = s^{(\rho_k \, \ldots \, \rho_1)}$ et que $\det s = s^{(1 \ldots 1)}$

Mesures fondamentales

dx désignera la mesure de Lebesgue associée aux structures euclidiennes définies par < . > (on n'oubliera pas que pour les matrices M associées à des vecteurs de R^{k-1}, $||M||^2 = < M, M > = < e, M \Delta M > = tr (M \Delta M) = 2 tr (M$

$= 2 \times (\text{norme usuelle de } M)^2$

$d\mu(x)$ désignera la mesure de Haar à gauche pour la loi $\perp$ qui est la mesure

$$x^{-(1 + \frac{k-1}{2})(1 \ldots 1)} dx$$

$d\nu(x)$ est la mesure de Haar à droite, la fonction modulaire étant

$$\frac{d\nu}{d\mu} = x^{(\alpha_1 \ldots \alpha_k)} \qquad \alpha_i = \frac{k+1}{2} - i$$

Présentation de (X, S) dans ces espaces homogènes

Nous avons vu que $X\,{}^tX$ était, pour une étude de statistique invariante, équivalente à X. Aussi, nous représenterons (X, S) sous la forme de la matrice (k+1, k+1) symétrique $\begin{pmatrix} S & X \\ {}^tX & 0 \end{pmatrix}$, ce qui permettra de faire jouer tout son rôle à la loi Δ. Les composantes S et X sont indépendant On posera ici $a^* = \frac{1}{4} \Lambda^{-1}$. La composante X a la densité

$$\frac{1}{\sqrt{2\pi}} \, \frac{e^{-<a^*, (x-m) \Delta (x-m)>}}{a^{1/2(1 \ldots 1)}}$$

par rapport à la mesure de Lebesgue de l'espace euclidien R^k muni de < , > . Enfin S suit la loi de Wishart à $\rho-1$ degrés de liberté et si $\rho > k$ admet la densité (par rapport à $d\mu(x)$)

$$\frac{e^{-<a^*, s>} \; s^{\frac{\rho-1}{2}(1,1\ldots 1)}}{\Gamma\left[\frac{\rho-1}{2}(1,1\ldots 1)\right] \; a^{\frac{\rho-1}{2}(1,1\ldots 1)}}$$

Pour le calcul du coefficient Γ , on se reportera à [5] .

IV- Un galop d'essai : le test de Hotelling

Nous développerons de façon détaillée cette solution du problème classiq qui consiste à tester $m = 0$ contre $m \neq 0$, bien que les résultats obtenus ne soient pas nouveaux. En effet c'est une bonne méthode pour s'initier aux techniques que nous développerons sur des problèmes plus complexes.

Le groupe linéaire est engendré par le groupe triangulaire et le groupe orthogonal. Notre attention se porte sur la statistique $(X\,{}^tX, S)$ ou mieux encore $(X \Delta X, S)$. Faire opérer une matrice triangulaire T sur les données revient à faire les produits à gauche par $b = T\,{}^tT$, c'est-à-dire à transformer $(X \Delta X, S)$ en $(b \perp (X \Delta X), b \perp S)$. Comme $b \perp (X \Delta X)$ est un point de la frontière du cône $\mathcal{M}_k^+$, on ne trouvera de statistique invariante que si $b \perp S$ est dans $\mathcal{M}_k^+$, et donc inversible (pour $\perp$). On est ramené à considérer la statistique invariante (par le groupe triangulaire)

$$(b \perp S)^{\perp -1} \perp b \perp (X \Delta X) = S^{\perp -1} \perp (X \Delta X)$$

définie si ρ est strictement supérieur à k, (ce que nous supposerons désormais). L'action du groupe orthogonal fait subir à cette dernière statistique des rotations (au sens du produit scalaire $<\ ,\ >$) autour de la matrice identité e de $\mathcal{M}_k^+$. D'où le choix de la statistique $T = < e\ ,\ S^{\perp -1} (X \Delta X >$ qui s'écrit encore $T = ||S^{-1} \perp (X)||^2$ (exercice : vérifier que T est la statistique de Hotteling, au coefficient ρ-1 près).

Théorème 1

La répartition de T est la loi β décentrée, de paramètre de décentrage. $\delta = \frac{1}{2} ||\Lambda^{-1} \perp (m)||^2 = {}^tm\ \Lambda^{-1}\ m$.

Nous donnerons seulement la méthode de démonstration de ce résultat , qui est cité dans [6] , les calculs explicites étant repris au chapitre suivant.

La fonction de répartition de T est donnée pour $t > 0$ par

$$P\,(T < t) = k\,(a) \iint_{||s^{-1} \perp (x)||^2 < t} e^{-\langle a^*, s + (x-m)\Delta(x-m)\rangle} \; s^{\frac{\rho-1}{2}(1\ldots1)} \, d\mu\,(s)\,dx$$

$$= k(a)\; e^{-||a \perp (m)||^2} \iint_{||s^{\perp -1} \perp(x)||^2 < t} e^{-\langle a^*, s + x\Delta x\rangle} \; e^{2\langle a^*, x\Delta m\rangle} \; s^{\frac{\rho-1}{2}(1\ldots1)} \, d\mu\,(s)\,dx$$

On effectue les changements de variables $a \perp s = s'$ qui conserve $d\mu\,(s)$, $a \perp (x) = x'$ (de jacobien $a^{1/2\,(1\ldots1)}$, cf [5]) puis le changement de variable (pour s fixé) $y = s^{\perp -1} \perp (x)$

Enfin l'écriture de y en coordonnées "polaires" $y = \sqrt{r}u$ $(||u|| = 1)$ met en évidence la densité de T exprimée en fonction de la mesure de Lebesgue $d\sigma\,(u)$ de la sphère unité (pour $\langle\, ,\, \rangle$) de R^k. Pour une matrice orthogonale H convenablement choisie le changement de variable, défini par

$$\begin{cases} s' = {}^tH\,s\,H \\ s' \perp (v) = H\,[s \perp (u)] \end{cases}$$

ramène le calcul au cas où $a \perp (m) = ||a \perp (m)|| \begin{bmatrix} 0 \\ \vdots \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

On développe alors la dernière exponentielle en série et on permute les signe $\sum$ et $\iint$. Les termes de rang impairs sont nuls par symétrie. Ceux de rang pair se calculent en écrivant $d\mu\,(s) = s^{-\alpha}\,d\nu\,(s)$ et en effectuant le changement de variable $s' = s \perp (e + t\,v\,\Delta\,v)$ qui permet d'intégrer en s' puis par des méthodes usuelles en v.

<u>Théorème 2</u>

Le test de Hotelling est UMP parmi les tests invariant sous le groupe linéaire

Preuve : il suffit d'appliquer au résultat précédent le lemme de Neymann-Pearson. Le rapport de vraisemblance s'écrit $\sum_{j \in N} a_j \left(\frac{t}{1+t}\right)^j$, les a_j étant des coefficients positifs.

C'est donc une fonction croissante de $\frac{t}{1+t}$, et a fortiori de t.

V- Problèmes où interviennent plusieurs échantillons

De nombreux problèmes se ramènent au problème théorique suivant. Soient $(\overline{X}_1, \ldots, \overline{X}_j, S)$ une statistique dont les composantes sont indépendantes et où les X_i sont des variables normales de R^k d'espérance m_i, de covariance $\frac{1}{\rho_i}\Lambda$ et où S suit une loi de Wishart de paramètre Λ à $\sum_{i=1}^{j}(\rho_i - 1)$ degrés de liberté. On se propose de tester l'hypothèse "tous les m_i sont nuls" contre des hypothèses adverses. Remarquons que sans de très fortes restrictions sur ces hypothèses, on n'obiendra pas de solution satisfaisante : contre une hypothèse où tous les m_i sont nuls sauf m_{i_0}, le meilleur test sera, en vertu du IV le test de Hotelling $||S^{\perp -1} \perp (X_{i_0})||^2 > t$ qui dépend de i_0. De plus en l'absence probable de test UMP, nous nous bornerons à rechercher des tests localement UMP (qui séparent le plus vite des hypothèses voisines).

Sur ce sujet, le lecteur consultera avec fruit, outre le classique ouvrage [1], l'intéressante étude [3] sur les tests usuellement pratiqués.

Posons de même qu'en IV $X_i = \sqrt{\frac{\rho_i}{2}}$ et reprenant la même méthode nous sommes conduit à étudier les statistiques $(S, X_i \Delta X_{i'})$ puis les statistiques $S^{\perp -1} \perp (X_i \Delta X_{i'})$ et enfin à nous borner aux distances mutuelles entre ces points, et entre ces points et l'origine.

VI- Une bonne approximation

Nous allons oubier dans un premier temps les statistiques "rectangles" et ne conserver que les statistiques $T_i = ||S^{\perp -1} \perp (X_i)||^2$.

Décrivons leur loi conjointe

Théorème 3

La densité de $(T_1\ T_2 \dots T_j)$ est proportionnelle à

$$e^{-\sum_{i=1}^{j} ||a\perp(m_i)||^2} \prod_{i=1}^{j} t_i^{\frac{k}{2}-1} \iint\dots\int e^{-<e,s\perp(e+\sum_{i=1}^{j} t_i\, u_i \Delta u_i)>}$$

$$e^{2\sum_{i=1}^{j} <a\perp(m_i),\ s\perp(u_i)> \sqrt{t_i}}\ s^{\frac{\rho}{2}(1,1,\dots 1)}\ d\mu(s) \prod_{i=1}^{j} d\sigma(u_i)$$

Preuve : la fonction de répartition conjointe des T_i est, en posant

$$k(a) = \frac{1}{(2\pi)^{\frac{j}{2}}\ a^{\frac{\rho}{2}(1\dots 1)}}$$

$$P\left(\bigcap_{i=1}^{j} (T_i < t_i)\right) = k(a) \iint\dots\int_{||s^{\perp-1}\perp(x_i)||^2 < t_i} e^{-<a^*,s+\sum_{i=1}^{j}(x_i\ \ m_i)\Delta(x_i-m_i)>}$$

$$s^{\frac{\rho-j}{2}(1\dots 1)}\ d\mu(s) \prod_{i=1}^{j} dx_i$$

On pose $a\perp s = s'$ et $a\perp(x) = x'$.

La mesure de Haar à gauche est conservée et dx_i est conservé au coefficient $a^{\frac{1}{2}(1\dots 1)}$ près. D'où :

$$P\left(\bigcap_{i=1}^{j}(T_i < t_i)\right) = k(e) \iint\dots\int_{||s^{\perp-1}\perp(x_i)||^2 < t_i} e^{-<e,s+\sum_{i=1}^{j}(x_i-a\ \ (m_i))\Delta(x_i-a\ \ (m_i))}$$

$$s^{\frac{\rho-j}{2}(1,1\dots 1)}\ d\mu\ (s) \prod_{i=1}^{j} dx_i$$

$$= k(e)\ e^{-\sum_{i=1}^{j}||a\perp(m_i)||^2} \iint\dots\int_{||s^{\perp-1}\perp(x_i)||^2 < t_i} e^{-<e,s+\sum_{i=1}^{j} x_i \Delta x_i>}$$

$$e^{2\sum_{i=1}^{j} <a\perp(m_i),\ x_i>}\ s^{\frac{\rho-j}{2}}\ d\mu(s) \prod_{i=1}^{j} dx_i$$

Pour s fixé on pose $x_i = s \perp(y_i)$ qui fait intervenir le jacobien $s^{\frac{1}{2}(1\ldots1)}$.

Alors

$$P\left(\bigcap_{i=1}^{j}(T_i<t_i)\right) = k(e)\; e^{-\sum_{i=1}^{j}||a\perp(m_i)||^2} \iint\cdots\int_{||y_i||^2<t_i} e^{-<e,s\perp(e+\sum_{i=1}^{j} y_i \,\Delta\, y_i)}$$

$$e^{2\sum_{i=1}^{j}<a\perp(m_i),\; s\perp(y_i)>}\; s^{\frac{\rho}{2}(1,1\ldots1)} \; d\mu(s) \prod_{i=1}^{j} dy_i$$

On pose enfin $u_i = \dfrac{y_i}{||y_i||}$ $\quad r_i = ||y_i||^2$

ce qui donne $dy_i = r_i^{\frac{k}{2}-1}\; d\sigma(u_i)\, dr_i$, $d\sigma(u_i)$ étant la mesure de Lebesgue sur la surface de la sphère unité $||u_i|| = 1$. La densité cherchée est mise en évidence par la formule obtenue :

$$P\left(\bigcap_{i=1}^{j}(T_i<t_i)\right) = k(e)\; e^{-\sum_{i=1}^{j}||a\perp(m_i)||^2} \iiint_{r_i<t_i} e^{-<e,s\perp(e+\sum_{i=1}^{j} r_i u_i \,\Delta\, u_i)>}$$

$$e^{2\sum_{i=1}^{j}<a\perp(m_i),\; s\perp(u_i)>\sqrt{r_i}}\; s^{\frac{\rho}{2}(1,1\ldots1)}\, d\mu(s) \prod_{i=1}^{j} r_i^{\frac{k}{2}-1}$$

$$\prod_{i=1}^{j} dr_i \prod_{i=1}^{j} d\sigma(u_i)$$

Théorème 4

Notons par $f(t_1\, t_2 \ldots t_j)$ l'intégrale exprimée au théorème 3 sous l'hypothèse nulle. Pour la structure statistique $(T_1\, T_2 \ldots T_j)$, il existe un test localement UMP de l'hypothèse $(\forall i \quad m_i=0)$ contre toute hypothèse de la forme $(||a\perp(m_1)|| = ||a\perp(m_2)|| = \ldots = ||a\perp(m_j)||$

Sa région de rejet est donnée par

$$- \sum_{i=1}^{j} t_i \frac{\partial f}{\partial t_i} \quad > c\, f$$

(N'oublions pas que $||a \perp (.)||$ définit la distance naturelle entre observations associées à l'expérience, car $||a \perp (m)||^2$ n'est autre que ${}^t m\, \Lambda^{-1}\, m$,.... même si cette distance est inconnue !)

Démonstration

Le paramètre est ici

$$m = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_j \end{pmatrix}$$

qui est un élément de R^{kj} .

Les tests localement UMP sont ceux qui (pour des tests sans biais de niveau donné) maximisent en m = 0 (hypothèse nulle) la forme quadratique des dérivées secondes. Nous écrivons la matrice de la différentielle seconde $D^2_{m,m} \quad [\mu, \mu]$ par ses blocs de dimension (k, k). Cela revient à calculer les différentielles partielles $D^2_{m_i\, m_{i'}} \quad [\mu_i\ ,\ \mu_{i'}]$ puisqu'alors

$$D^2_{mm} = \left(\begin{array}{c|c|c} D^2_{m_1 m_1} & & \\ \hline & & \\ & & \boxed{D^2_{m_i\, m_{i'}}} \end{array} \right)$$

Nous ferons m = 0 dans cette expression. On a d'abord, pour la densité $g\ (t_1 \ldots t_j)$

$$D_{m_i} g \; [\mu_i] = 2 < a \perp (m_i), a \perp (\mu_i) > g \; (t_1 \ldots t_j)$$

$$+ e^{-\sum_{i=1}^{j} ||a \perp (m_i)||^2} \prod_{i=1}^{j} t_i^{\frac{k}{2}-1} \iint \ldots \int e^{-<e, s \perp (e + \sum_{i=1}^{j} t_i u_i \Delta u_i)>}$$

$$2 \sqrt{t_i} < a \perp (\mu_i), s \perp (u_i) > e^{2 \sum_{i=1}^{j} <a \perp (m_i), s \perp (u_i)> \sqrt{t_i}}$$

$$s^{\frac{\rho}{2}(1 \ldots 1)} \; d\mu \; (s) \; \prod_{i=1}^{j} d\sigma \; (u_i)$$

On calcule ensuite pour $i \neq i'$ $D_{m_i \, m_{i'}} \left[\mu_i \, , \, \mu_{i'}\right]$

Il est facile de voir que pour $m_i = m_{i'} = 0$, cette forme est identiquement nulle, les intégrales étant nulles par des arguments de symétrie. Il n'en est pas de même pour $i = i'$.

En $m_i = 0$, la différentielle seconde partielle $D^2_{m_i, m_i} \left[\mu_i, \mu_i\right]$ s'exprime par

$$D^2_{m_i m_i} \left[\mu_i, \mu_i\right] = 2 \; ||a \perp (\mu_i)||^2 \prod_{i=1}^{j} t_i^{\frac{k}{2}-1} \quad f \; (t_1 \ldots t_j)$$

$$+ \prod_{i=1}^{j} t_i^{\frac{k}{2}-1} \iint \ldots \int e^{- <e, s \perp (e + \sum_{i=1}^{j} (e + t_i u_i \Delta u_i)>}$$

$$4 \; t_i \; (< a \perp (\mu_i), s \perp (u_i) >)^2 \; s^{\frac{\rho}{2}(1 \ldots 1)} \; d\mu(s) \prod_{i=1}^{j} d\sigma(u_i)$$

Remarquons ensuite que $< a \perp (\mu_i), s \perp (u_i) >$ est la trace de

$2 \left[a \perp (\mu_i)\right] \; {}^t\left[s \perp (u_i)\right]$ et vaut donc $2 \; {}^t\left[s \perp (u_i)\right] \; \left[a \perp (\mu_i)\right]$.

Son carré s'exprime donc comme $4 \; {}^t \left[a \perp (\mu_i)\right] \left[s \perp (u_i)\right]^t \left[s \perp (u_i)\right] \left[a \perp (\mu_i)\right]$

ou encore $2 \; {}^t \left[a \perp (\mu_i)\right] \; \left[s \perp (u_i) \Delta \; s \perp (u_i)\right] \; \left[a \perp (\mu_i)\right]$

ou enfin $2 \; {}^t \left[a \perp (\mu_i)\right] \; \left[s \perp (u_i \Delta u_i)\right] \; \left[a \perp (\mu_i)\right]$.

On écrira alors :

$$D^2_{m_i m_i}\left[\mu_i,\mu_i\right] = 4\ {}^t\!\left[a \perp(\mu_i)\right]\ \left[f\ (t_1\ldots t_j)\ e + M_i\right]\ \left[a \perp(\mu_i)\right]$$

en posant :

$$M_i = 4\ t_i \iint\cdots\int e^{-<e,s\perp(e + \sum_{i=1}^{j} t_i\ (u_i \Delta u_i))>}$$

$$\left[s \perp(u_i)\right]\ {}^t\!\left[s \perp(u_i)\right]\ s^{\frac{\rho}{2}(1\ldots1)}\ d\mu(s) \prod_{i=1}^{j} d\sigma\ (u_i)$$

Calculons maintenant ${}^tH\ M_i\ H$, pour H matrice orthogonale.

$${}^tH\ M_i\ H = 4\ t_i \iint\cdots\int e^{-<e,s\perp(e + \sum_{i=1}^{j} t_i\ (u_i\ \Delta\ u_i))>}$$

$${}^tH\ \left[s \perp(u_i)\right]^t \left[s \perp(u_i)\right]\ H\ s^{\frac{\rho}{2}(s\ldots1}\ d\mu(s) \prod_{i=1}^{j} d\sigma(u_i)$$

Effectuons le changement de variable défini par

$$\begin{cases} s' = {}^tH\ s\ H \\ s' \perp(v_i) = {}^tH\left[s \perp(u_i)\right] \end{cases}$$

On obtient alors les relations

$$< e,\ s > = < e,\ s' >$$

$$< e,\ s \perp(u_i\ \Delta\ u_i) > = 2\ {}^t\left[s \perp(u_i)\right]\ \left[s \perp(u_i)\right]$$

$$= 2\ {}^t\left[s' \perp(v_i)\right]\ {}^tH\ \ H\left[s' \perp(v_i)\right] = < e,\ s' \perp(v_i\ \Delta\ v_i) >$$

Opérant sur la variable s, ce changement de variable conserve l'orthogonalit

donc la mesure ds.

Comme $d\mu\ (s) = \left[\det\ (s)\right]^{-\frac{k+1}{2}}$ et que $s^{(1\ldots1)} = \det s$ il conserve aussi

$s^{\frac{\rho}{2}(1\ldots1)}$ $d\mu\ (s)$. Si de plus pour s fixé, $u_1(z)$ et $u_2(z)$ sont deux courbe

orthogonales en $z = 0$ tracées sur la sphère $||u|| = 1$, c'est-à-dire vérifian

$< u_1'\ (0),\ u_2'\ (0) > = 0$, pour les courbes homologues définies par

$$v_i = s'^{\perp-1} \perp({}^tH\ s \perp(u_i))\ \text{on a :}$$

$$v'_i = s'^{\perp -1} \perp ({}^tH\, s \perp (u'_i))$$

Posant $s = T\, {}^tT$ et $s' = T'\, {}^tT'$, où T et T' sont des matrices triangulaires supérieures, cette relation s'écrit $v'_i = T'^{-1}\ {}^tH\, T\, u'_i$ et par conséquent

$$< v'_1(0),\ v'_2(0) > = 2\ {}^t u'_1(0)\ {}^tT\ M\ {}^tT'^{-1}\ T'^{-1}\ {}^tH\ T\ u'_2(0)$$

$$= 2\ {}^t u'_1(0)\ {}^t\ T\ H\ s'^{-1}\ {}^tH\ T\ u'_2(0)$$

$$= 2\ {}^t u'_1(0)\ {}^tT\ s^{-1}\ T\ u'_2(0)$$

$$= 2\ {}^t u'_1(0)\ {}^tT\ {}^tT^{-1}\ T^{-1}\ T\ u'_2(0)$$

$$= < u'_1(0),\ u'_2(0) > = 0$$

Aussi, pour s fixé, les changements de variables en u_i conservent les $d\sigma\ (u_i)$. Comme la première expression du changement de variable global ne fait pas intervenir u, il s'ensuit que

$$s^{\frac{\rho}{2}(1 \ldots 1)}\ d\mu\ (s)\ \prod_{i=1}^{j}\ d\sigma\ (u_i)$$ est conservé, donc que l'intégrale est

inchangée. On a donc, pour toute matrice H orthogonale ${}^tH\ M_i\ H = M_i$, ce qui prouve que M_i s'écrit $u_i\ (t_1, \ldots t_j)\ e$,

avec $u_i\ (t_1 \ldots t_j) = \dfrac{< e,\ M_i >}{||e||^2}$

On en déduit :

$$u_i\ (t_1 \ldots t_j) = k\ t_i \iint \ldots \int e^{- <e,\ s \perp (e + \sum_{i=1}^{j} t_i\, u_i\, \Delta\, u_i) >}$$

$$< e,\ s \perp (u_i\ \Delta\ u_i) > s^{\frac{\rho}{2}(1..1)}\ d\mu(s)\ M\ d\sigma(u_i)$$

$$= - k\ t_i \frac{\partial f}{\partial t_i}$$

Le lemme de Neyman Pearson conduit donc, pour maximiser la forme différentiell seconde, à prendre comme région de rejet

$$\sum_{i=1}^{j} ||a \perp (\mu_i)||^2\ (f - k_1\ t_i \frac{\partial f}{\partial t_i})\ >\ c_1\ f$$

ou plus simplement

$$- \sum_{i=1}^{j} ||a \quad (\mu_i)||^2 \quad t_i \frac{\partial f}{\partial t_i} > c_2 f$$

Si on se restreint à des hypothèses adverses vérifiant

$||a \perp (\mu_1)|| = ||a \perp (\mu_2)|| = \dots = ||a \perp (\mu_j)||$, cela revient à rejeter si

$$- \sum_{i=1}^{j} t_i \frac{\partial f}{\partial t_i} > c f$$

<u>Théorème 5</u> (Principe de l'analyse de la variance unidimentionnelle)

Si les statistiques X_i sont à valeurs dans R, le test ci-dessus rejette l'hypothèse nulle si $\sum_{i=1}^{j} T_i > C$

<u>Preuve</u> . $d\sigma$ s'écrit simplement $\delta_1 + \delta_{-1}$ (δ_x mesure de Dirac en x) et $u_i \Delta u_i = u_i^2$; enfin $d\mu$ (s) est la mesure $\frac{ds}{s}$. On en déduit

$$f(t_1 \dots t_j) = K \int e^{-s\,(1 + \sum_{i=1}^{j} t_i)} \; s^{\frac{\rho}{2} - 1} \, ds$$

$$= \frac{K \; \Gamma\left(\frac{\rho}{2}\right)}{\left(1 + \sum_{i=1}^{j} t_i\right)^{\frac{\rho}{2}}}$$

La région de rejet est donc donnée par

$$\frac{\sum_{i=1}^{j} t_i}{1 + \sum_{i=1}^{j} t_i} > \frac{2\,c}{\rho}$$

ou encore par $\sum_{i=1}^{j} t_i > c'$ puisque $\frac{x}{1+x}$ est croissante pour $x > 0$.

<u>Théorème 6</u>

Le théorème 5 n'est plus valable pour $k > 1$. Cependant si $\rho \to \infty$, le test construit au théorème 4 est asymptotiquement équivalent au test de Lawley-Hotteling $\sum_{i=1}^{j} T_i > c$

Preuve. $k = 2$, $j = 2$ et $\rho = 4$ fournissent un contre exemple accessible par un calcul élémentaire de $f\ (t_1, t_2)$

Posant enfin $R_i = \rho\ T_i$, le test précédent rejette donc si

$$-\sum_{i=1}^{j} \frac{R_i}{\rho} \frac{\partial f}{\partial t_i} \left(\frac{R_1}{\rho}, \frac{R_2}{\rho} \dots \frac{R_j}{\rho}\right) > cf \left(\frac{R_1}{\rho}, \frac{R_2}{\rho} \dots \frac{R_j}{\rho}\right)$$

Pour des v.a. indépendantes S (ayant la répartition de Wishart à ρ degrés de liberté) et U_i (uniformes sur la sphère $||U_i|| = 1$) on-interprète f et $\frac{\partial f}{\partial t_i}$ comme

$$\begin{cases} f\left(\frac{r_1}{\rho}, \frac{r_2}{\rho} \dots \frac{r_j}{\rho}\right) = E\left(e^{-<e,\ \frac{S}{\rho} \perp \left(\sum_{i=1}^{j} r_i\ U_i\ \Delta\ u_i\right)>}\right) \\ -\frac{1}{\rho} \frac{\partial f}{\partial t_i} \left(\frac{r_1}{\rho}, \frac{r_2}{\rho} \dots \frac{r_j}{\rho}\right) = E\left(e^{-<e,\ \frac{S}{\rho} \perp \left(\sum_{i=1}^{j} r_i\ U_i\ \Delta\ U_i\right)>} < e, \frac{S}{\rho} \perp (U_i \Delta U_i)>\right. \end{cases}$$

Si $\rho \to \infty$ $\frac{S}{\rho} \to e$ en loi (loi des grands nombres),

et $f\left(\frac{r_1}{\rho} \dots \frac{r_j}{\rho}\right) \to E\left(e^{-\sum_{i=1}^{j} r_i\ ||u_i||^2}\right) = e^{-\sum_{i=1}^{j} r_i}$

De même

$$-\frac{1}{\rho} \frac{\partial f}{\partial t_i} \left(\frac{r_1}{\rho} \dots \frac{r_j}{\rho}\right) \to e^{-\sum_{i=1}^{j} r_i}$$

L'ensemble de rejet tend donc vers $\sum_{i=1}^{j} R_i > c$

VII- Ebauche d'étude générale

Dans le cas unidimensionnel les statistiques $S^{-1} \perp (X_i\ \Delta\ X_{i'})$ (c'est-à-dire plus simplement $\frac{X_i\ X_{i'}}{S}$ ... !) sont les seules statistiques invariantes sous le groupe linéaire (la seule transformation orthogonale non triviale est $X \to -X$ déjà utilisée). Or leur donnée est équivalente à celle des $T'_i = \frac{X_1\ X_i}{S}$.

La méthode précédente conduit au résultat négatif suivant :

Théorème 7

Il n'existe pas de test localement UMP de l'hypothèse nulle ($m_i = 0 \quad \forall i$ contre $|m_1| = |m_2| = \dots = |m_j|$

Preuve. un calcul classique montre que la densité conjointe des T'_i est :

$$k\, t_1^{-\frac{j}{2}}\, e^{-\sum_{i=1}^{j} \frac{m_i^2}{2\sigma^2}} \int_{R^+} e^{-s\,(1+\sum_{i=1}^{j} \frac{t_i^2}{t_1})}\, \mathrm{Ch}\left(\frac{\sqrt{s}}{\sigma} \sum_{i=1}^{t} \frac{m_i t_i}{\sqrt{t_1}}\right.$$

$$s^{\frac{\rho}{2}-1}\, ds$$

En différenciant deux fois, on voit que la zone de rejet d'un test localement UMP serait donnée par

$$\sum_{i=1}^{j} \sum_{i'=1}^{j} \frac{\mu_i\, \mu_{i'}\, T'_i\, T'_{i'}}{T'_1} > c\,(1+\sum_{i=1}^{j} \frac{T'^2_i}{T'_1})$$

ce test dépend donc étroitement du choix des signes des μ_i , d'où le résultat

Théorème 8

Dans une optique Bayésienne, si les m_i ont des répartitions symétriques autour de 0 et sont indépendantes, le test d'analyse de la variance est localement UMP contre les hypothèses $|m_1| = |m_2| = \dots = |m_j|$

Preuve. La répartition de m_i conditionnelle à $|m_i|$ est la loi $\frac{1}{2}\ \delta_{-|m_i|} + \frac{1}{2}\ \delta_{|m_i|}$. Par suite de l'indépendance, on a alors :

$$\begin{aligned} E\,(\mu_i\, \mu_{i'} \mid\ |\mu_1|\,,\, |\mu_2|\,,\, \dots |\mu_j|) &= 0 \quad \text{si } i \neq i' \\ &= \mu_i^2 \quad \text{si } i = i' \end{aligned}$$

Le calcul précédent conduit à la zone de rejet

$$E\left(\sum_{i=1}^{j} \sum_{i'=1}^{j} \frac{\mu_i\, \mu_{i'}\, t'_i\, t'_{i'}}{t'_1}\right) > c \left(1+\sum_{i=1}^{j} \frac{t'^2_i}{t'_1}\right)$$

ou après avoir conditionné par la connaissance des μ_i

$$E \left(\sum_{i=1}^{j} \mu_i^2 \frac{t_i'^2}{t_j'} \right) > c \left(1 + \sum_{i=1}^{j} \frac{t_i'^2}{t_1'} \right)$$

En imposant que tous les μ_i soient égaux en valeur absolue on est conduit à rejeter l'hypothèse nulle si $\sum_{i=1}^{j} \frac{T_i'^2}{T_1'} > c_1$ c'est-à-dire si $\sum_{i=1}^{j} T_i > c_1$

IX- Conclusion

Cette étude met en évidence des propriétés optimales des méthodes de l'analyse de la variance pour des études unidimensionnelles. En analyse multivariée, il est raisonnable de penser que le théorème 7 reste valable et qu'une variante du théorème 8 conduit à prendre les tests plus sophistiqués établis au théorème 4, dont le test de Lawley-Hotteling est la forme limite. Du point de vue pratique, ce dernier conduit à des bornes fonctions puissances et surtout à des calculs simples. Seules des simulations peuvent permettre de décider s'il est vraiment utile (pour des nombres de degrés de liberté faibles) de prendre des formes plus élaborées.

X- Bibliographie

[1] ANDERSON T.W. - An introduction to Multivariate Statistical Analysis Wiley, New-York (1958)

[2] BARRA J.R. - Notions fondamentales de statistique mathématique. Dunod - Paris (1971)

[3] EATON M.L. et PERLMAN M.D. - A monotoqicity property of the power functions of some invariant tests for MANOVA. Ann. Math. Statist. Vol. 2, N° 5 (1022-1028)

[4] FOURT G. - Lois sur les cônes homogènes. Annales Scientifiques de l'Université de Clermont-Ferrand, N° 51 (1974), p. 22-30

[5] FOURT G. - Mesures sur les cônes homogènes (soumis pour publication)

[6] LINNIK Y. - Leçons sur les problèmes de statistique analytique. Gauthier-Villars (1967)

[7] SCHEFFE H. - The analysis of variance. Wiley, New-York

[8] VINDBERG E.B. - Theory of convex homogenenous cones. Trans. Moscow Math. Society 12 (1963)

G. FOURT
Université de Clermont II
Complexe Scientifique des Cézeaux
Département de Mathématiques Appliquées
Boîte Postale n° 45
F- 63170 AUBIERE