Statistique et analyse des données

J. L. PHILOCHE

Une condition de validité pour le test F

Statistique et analyse des données, tome 2, nº 1 (1977), p. 37-59
<http://www.numdam.org/item?id=SAD_1977__2_1_37_0>

UNE CONDITION DE VALIDITE

POUR LE TEST F

par

J.L. PHILOCHE (1)

INTRODUCTION

On se propose dans un exposé autonome de montrer que le test F classique du modèle linéaire est exactement valide dans bien d'autres cas que l'usuel contexte Gaussien. On commence (§1) par énoncer tous les résultats probabilistes nécessaires à l'utilisation du type de lois de probabilité concernées. Ensuite (§2), après avoir resituer le cadre du modèle linéaire, on énonce le résultat statistique principal, on en précise l'intérêt et les limites, et on situe ce travail par rapport à la littérature existante. Enfin (§3) on donne les démonstrations de tous les énoncés présentés.

(1) Ecole Nationale Supérieure Agronomique de Rennes et Ecole Polytechnique, Palaiseau.

1. LOIS DE PROBABILITE SPHERIQUEMENT, OU ISOTROPIQUEMENT DISTRIBUEES DANS R^n.

Pour n entier ($n \geqslant 1$), on désigne par S_{n-1} la sphère unité de R^n (1) : $S_{n-1} = \{x \; ; \; x \in R^n \text{ et } ||x|| = 1\}$, par $O(n)$ le groupe orthogonal de R^n, c'est-à-dire le groupe constitué des transformations orthogonales de R^n.

Lemme 1

Il existe, sur S_{n-1}, une unique mesure de probabilité σ_{n-1} invariante par $O(n)$.

(démonstration au §3).

Autrement dit, il existe, sur (la tribu borélienne de) S_{n-1} une unique probabilité σ_{n-1}, telle que, pour tout $A \in O(n)$, l'image $A(\sigma_{n-1})$ de σ_{n-1} par A soit identique à σ_{n-1} ; cette invariance, par les rotations, permet légitimement d'appeler σ_{n-1} "la" probabilité uniforme sur S_{n-1}.

Définition 1

Soit $Y = (Y_k \; ; \; 1 \leqslant k \leqslant n)$ un vecteur aléatoire à valeur dans R^n, on dira que Y est sphériquement distribué (2), si sa loi P_Y est invariante par $O(n)$; on dira que Y est sphériquement distribué autour de $m = (m_k \; ; \; 1 \leqslant k \leqslant n)$, $m \in R^n$, si $Y-m$ est sphériquement distribué.

A ceci près qu'un vecteur sphériquement distribué peut (éventuellement) prendre la valeur 0 avec une probabilité positive, les vecteurs sphériquement distribués appartiennent au cadre plus général de ce que nous appellerons les vecteurs

(1) R^n est implicitement muni de son produit scalaire usuel : tout se transpose aussitôt au cas où un autre produit scalaire serait choisi.

(2) On trouve aussi dans la littérature l'expression synonyme "radialement distribué".

isotropiquement distribués [1].

Définition 2

Soit $Y = (Y_k \; ; \; 1 \leqslant k \leqslant n)$ un vecteur aléatoire à valeur dans R^n, on dira que Y est isotropiquement distribué si

(i) $Y \neq 0$, presque sûrement,

(ii) la loi de probabilité de $Y/||Y||$ est σ_{n-1} ;

On dira que Y est isotropiquement distribué autour de m, ($m \in R^n$), si Y-m est isotropiquement distribué.

Le particularisme des vecteurs sphériquement distribués est précisé par le résultat suivant :

Lemme 2

Soit Y un vecteur aléatoire à valeur dans R^n, tel que $Y \neq 0$ presque sûrement ; les propriétés (j) et (jj) suivantes sont équivalentes :

(j) Y est sphériquement distribué,

(jj) Y est isotropiquement distribué et il y a indépendance entre $Y/||Y||$ et $||Y||$.

(démonstration au §3).

On voit ainsi que la loi d'un vecteur aléatoire Y ($\neq$ 0,p.s) sphériquement distribué est entièrement connue dès que l'on se donne la loi de $||Y||$; la situation est tout autre pour la loi d'un vecteur aléatoire Y isotropiquement distribué : seule est

(1) Il ne semble pas qu'il existe dans la littérature de terme consacré ; l'appellation choisie , assez simple, n'est sans doute pas très précise, peut-être aurait-il été plus juste d'employer le néologisme "iso-angulairement distribué"...

imposée l'uniformité de la loi de l'angle polaire $Y/||Y||$; toute famille (mesurable) $(P_\theta\ ;\ \theta \in S_{n-1})$ de probabilités sur $(R^+)^*$ fournit une version possible des lois conditionnelles.

$$\{P\,[\,||Y|| \in \cdot \mid Y/||Y|| = \theta\,]\ ;\ \theta \in S_{n-1}\}\ .$$

Avant de poursuivre nous donnerons quelques exemples :

Ex. 1 (loi de Gauss sphérique sur R^n).

Désignant par dy la mesure de Lebesgue sur R^n, ce cas correspond à la situation où la loi P_Y de Y est

$$(1.1)\qquad P_Y(dy) = (2\Pi)^{-n/2}\ \exp\left[-\frac{1}{2\sigma^2}\,||y-m||^2\right]\sigma^{-n}\,dy\ ;\ \text{où}$$

σ^2 et $m = (m_k\ ;\ 1 \leqslant k \leqslant n)$ sont des paramètres ($\sigma^2 > 0$, $m \in R^n$).

Dans cet exemple, le vecteur Y est sphériquement distribué autour de m ; la situation correspond au cas où les diverses composantes Y_k de Y suivent indépendamment des lois de Gauss (scalaire), admettant m_k pour moyenne et σ^2 comme variance (commune).

Ex. 2 (loi de Cauchy sphérique sur R^n).

Ce cas correspond, cette fois-ci, à la situation où la loi P_Y de Y est

$$(1.2)\qquad P_Y(dy) = K_n\ \frac{a^{-n}\ dx}{(1+||y-m||^2/a^2)^{\frac{n+1}{2}}}\ ,\ \text{où}$$

a et $m = (m_k\ ;\ 1 \leqslant k \leqslant n)$ sont des paramètres ($a > 0$ et $m \in R^n$) et K_n une constante de normalisation.

La raison pour laquelle il est légitime d'appeler cette probabilité une loi de Cauchy multidimensionnelle est la suivante : on peut vérifier que pour tout $u = (u_k\ ;\ 1 \leqslant k \leqslant n) \in R^n$, la loi de $(u,y) = \sum_{k=1}^{n} u_k Y_k$ est une loi de Cauchy unidimensionnelle [1] admettant $(u,m) = \sum_{k=1}^{n} u_k m_k$ pour médiane et dont l'intervalle interquartile est $2a||u||$; l'exemple 2 fournit un exemple où le vecteur Y est sphériquement distribué autour de son vecteur des médianes m.

Pour plus de détails sur cette loi de Cauchy multidimensionnelle consulter [1].

Le dernier exemple montre, dans R^2, un cas de loi isotropiquement (mais non sphériquement) distribuée autour d'un paramètre $m = (m_1, m_2) \in R^2$.

<u>Ex. 3</u>.

Pour une constante positive K, prenons pour loi P_Y du vecteur $Y = (Y_1, Y_2)$ la probabilité ;

(1.3)
$$P_{Y_1, Y_2}(dy_1, dy_2) = \frac{1}{4\pi K} \exp\left[-\frac{1}{K}|(y_1-m_1)(y_2-m_2)|\right] \frac{|(y_1-m_1)(y_2-m_2)|\,dy_1\,dy_2}{(y_1-m_1)^2 + (y_2-m_2)^2}$$

On peut vérifier que, dans cet exemple, le vecteur Y admet $m = (m_1, m_2)$ comme vecteur des moyennes, mais que ni Y_1, ni Y_2 ne possèdent une variance finie.

Les lois de probabilité distribuées sphériquement, ou isotropiquement, dans R^n conservent cette propriété par projectivité :

(1) Sur R, la loi de Cauchy est la probabilité $\frac{1}{\pi} \frac{dx/b}{1+(x-\mu)^2/b^2}$ elle ne possède pas de moyenne, toutefois μ est sa médiane, tandis que $\mu-b$ et $\mu+b$ sont respectivement le premier et le troisième quartile.

Lemme 3

Soit Y un vecteur aléatoire à valeur dans R^n, de loi P_Y, Π un projecteur orthogonal de R^n dont l'image est notée Im Π, alors :

(k) Si Y est sphériquement distribué dans R^n, il en est de même de ΠY dans Im Π ;

(kk) Si Y est isotropiquement distribué dans R^n, il en est de même de ΠY dans Im Π, dès que rg $\Pi \geqslant 1$.
Im Π est, naturellement, muni du produit scalaire induit.
(démonstration au §3).

Le vecteur m ($m \in R^n$) autour duquel un vecteur aléatoire Y est isotropiquement distribué est nécessairement "le" vecteur des médianes de Y, en effet :

Lemme 4

Soit $Y = (Y_k\ ;\ 1 \leqslant k \leqslant n)$ un vecteur aléatoire isotropiquement distribué dans R^n ($n \geqslant 2$), 0 est médiane unique de chaque composante Y_k (1).
(démonstration au §3).

d_1, d_2 désignant des entiers supérieurs ou égaux à 1, on rappelle que la loi de Fisher-Snedécor $F(d_1, d_2)$ est la probabilité sur R^+ définie par

(1.4) $$C(d_1 d_2)\ \frac{v^{d_1/2\ -1}}{\left[1 + \frac{d_1}{d_2}\, v\right]^{(d_1+d_2)/2}}\, dv\ ,$$

où $C(d_1, d_2)$ est une constante de normalisation convenable ; le résultat suivant est essentiel pour la suite :

(1) On perd manifestement l'unicité pour n = 1.

Lemme 5

Soit Y un vecteur aléatoire isotropiquement distribué dans R^n ; soient Π_1 et Π_2 deux projecteurs orthogonaux et orthogonaux entre eux dans R^n, de rangs respectifs d_1 et d_2, dans ces conditions, la variable aléatoire

$$\frac{||\Pi_1 Y||^2/d_1}{||\Pi_2 Y||^2/d_2} \text{ suit la loi Fisher-Snédécor } F(d_1, d_2)$$

(démonstration au §3).

Enfin, on peut établir que le cas gaussien (exemple 1) joue dans tout ceci un rôle très particulier :

Lemme 6

Soit $Y = (Y_k\ ;\ 1 \leq k \leq n)$ un vecteur aléatoire à valeur dans $R^n (n \geq 2)$; on suppose que Y est sphériquement distribué et que les coordonnées de Y sont indépendantes, alors les Y_k possèdent une même distribution gaussienne centrée (1).

(démonstration au §3).

2. TEST F ET VECTEUR ALEATOIRE ISOTROPIQUEMENT DISTRIBUE DANS R^n AUTOUR DU VECTEUR DES MEDIANES.

On désigne par $Y = (Y_k\ ;\ 1 \leq k \leq n)$ et $\varepsilon = (\varepsilon_k\ ;\ 1 \leq k \leq n)$ des vecteurs aléatoires à valeur dans R^n, par $\beta = (\beta_j ; 1 \leq j \leq p)$ un élément de R^p, par X (2) une application linéaire de R^p dans R^n (3). On suppose que

$$(2.1) \qquad Y = X(\beta) + \varepsilon \quad ,$$

(1) le résultat est tout à fait faux pour $n = 1$.

(2) Cette notation, classique, paraît assurément malheureuse au probabiliste ignorant le contexte des applications dont elle est issue (régression linéaire).

(3) n et p sont des entiers supérieurs ou égaux à 1, on suppose $p < n$.

et que seul Y est observable, le vecteur d'écart ε est en quelque sorte un élément caché. On suppose que X est connue mais que β ne l'est pas. Faisant certaines hypothèses "a priori" sur la loi de probabilité P_ε de ε, divers problèmes statistiques surgissent, en particulier celui du test d'une hypothèse linéaire auquel nous nous limiterons ci-dessous :

Il s'agit, au vu d'une réalisation y de Y de procéder au test d'une hypothèse linéaire concernant β, hypothèse selon laquelle β appartient à un sous espace vectoriel (s.e.v.) donné H_0 de R^p.

On remarquera que, sauf à préciser les "a priori" sur P_ε, le cadre précédent est exactement celui du modèle linéaire classique, cadre dans lequel on peut, en particulier, insérer analyse de variance et régression linéaire.

Tant que n (le nombre d'observations) ne varie pas, il est usuel de tout transférer dans R^n : en effet, dans R^n, le point $m = X(\beta)$ est, "a priori", susceptible de varier dans l'image L de X, tandis que l'exactitude de l'hypothèse à tester limiterait m à varier dans $L_0 = X(H_0)$; on désignera par Π et Π_0 les projecteurs orthogonaux sur L et L_0, on notera $\ell = \dim L$ et $\ell_0 = \dim L_0$ (on suppose $0 \leqslant \ell_0 < \ell \leqslant p$).

Nous en venons maintenant au résultat statistique essentiel :

<u>Proposition</u>

Soit $Y = (Y_k\ ;\ 1 < k < n)$ un vecteur aléatoire à valeur dans R^n, <u>isotropiquement distribué autour de son vecteur des médianes</u> $m = (m_k\ ;\ 1 \leqslant k \leqslant n)$, on suppose que m appartient à un s.e.v. L de R^n.

Soit L_0 un s.e.v. de L, F la variable aléatoire définie par

(2.2) $$F = \frac{||\Pi Y - \Pi_0 Y||^2/(\ell - \ell_0)}{||Y - \Pi Y||^2/(n-\ell)} \quad ;$$

Quand $m \in L_0$, la loi de la variable aléatoire F est la loi de Fisher-Snédécor $F(\ell - \ell_0, n-\ell)$

(démonstration au §3).

L'intérêt statistique du résultat précédent est clair si l'on désire au vu d'une réalisation de Y procéder au test de l'appartenance de m à L_0, considérant "a priori" que le vecteur des médianes m est inconnu dans L : pour tout $c > 0$, la région $w_c = [F > c]$ est la région critique d'un test dont le niveau est constant quand l'hypothèse nulle $m \in L_0$ est satisfaite, il ne reste plus qu'à déterminer c en fonction du seuil de risque que l'on s'est fixé à l'aide des tables de la distribution de Fisher-Snédécor.

Les hypothèses de la proposition sont en particulier satisfaites quand (exemple 1 du §1) la loi de Y est une loi de Gauss sphériquement distribuée autour de m ; ce cas est naturellement parfaitement bien connu, on rappelle qu'alors les coordonnées Y_k sont indépendantes et de même variance et que le vecteur des médianes est aussi le vecteur des moyennes. Dans ce cadre, la fonction puissance du test F qui s'exprime à l'aide de la loi de Fisher-Snédécor décentrée possède plusieurs propriétés d'optimalité remarquables, cf [2] et [3].

Les hypothèses de la proposition sont aussi satisfaites (exemple 2 du §1) quand la loi de Y est une loi de Cauchy sphériquement distribuée autour de son vecteur des médianes ; cette situation apparaîtra sans doute comme plus surprenante aux usagers du modèle linéaire habitués à des hypothèses concernant les moments du vecteur d'écart ε.

Enfin, l'exemple 3 du §1, donne un cas, peut-être un peu académique, mais simple, où les hypothèses sont à nouveau satisfaites sans pour autant que la loi de Y soit sphériquement distribuée autour de m.

Il n'est sans doute pas inutile de s'apercevoir que le résultat précédent peut être interprété comme un résultat particulièrement bon sur la <u>robustesse</u> (1) du test F autour de la situation gaussienne qui consiste en la donnée,

m_i) d'un modèle statistique initial :
la loi P_Y est inconnue dans la famille $\mathcal{G}_L$ des lois de Gauss sphériquement distribuées autour d'un vecteur des moyennes censé appartenir à un s.e.v. L de R^n,

p_i) d'un problème statistique initial :
pour $P_Y \in \mathcal{G}_L$, désignant par $g(P_Y)$ le vecteur des moyennes de P_Y, on se demande si l'on peut admettre l'hypothèse nulle h_0 définie par :

$$P_Y \in h_0 \text{ si et seulement si } P_Y \in \mathcal{G}_L \text{ et } g(P_Y) \in L_0 ,$$

s_i) d'une solution à ce problème initial :
le test $[F > c]$, pour un point critique c bien choisi.

Or cette situation gaussienne initiale peut être immergée dans une situation élargie consistant en la donnée,

m_e) d'un modèle statistique élargi :
la loi P_Y est inconnue dans la famille $\mathcal{S}_L$ des probabilités sur R^n isotropiquement distribuées autour d'un vecteur de médianes appartenant à L ;
on observera que $\mathcal{S}_L \supset \mathcal{G}_L$,

(1) L'approche que nous avons ici de la robustesse doit beaucoup à des idées proposées par Ph. COURREGE vers 1970.

p_e) d'un problème statistique élargi :
pour $P_Y \in \mathcal{P}_L$, désignant par $\hat{g}(P_Y)$ le vecteur des médianes de P_Y, on se demande si l'on peut admettre l'hypothèse nulle H_0 définie par :

$P_Y \in H_0$ si et seulement si $P_Y \in \mathcal{P}_L$ et $\hat{g}(P_Y) \in L_0$;
on observera que la fonction $\hat{g}$ est un prolongement de g à $\mathcal{P}_L$ et que H_0 est un "prolongement" de h_0 à $\mathcal{P}_L$, puisque $H_0 \cap \mathcal{G}_L = h_0$.

Le problème de robustesse consiste à apprécier en quel sens la solution du problème initial est une solution "honorable" du problème élargi : dans le cas qui nous occupe la robustesse du test F est parfaite puisque le niveau, constant, du test ne change pas de la situation initiale à la situation élargie !

En plus des lignes précédentes, s'intéressant à la robustesse du test F, il paraît souhaitable d'ajouter quelques remarques sur la portée du résultat présenté.

Dans le cadre présenté, nous trouvons intéressant de pouvoir utiliser le test F usuel pour apprécier l'exactitude d'une hypothèse linéaire concernant les médianes d'un vecteur aléatoire. Il est même assez spectaculaire que le cadre proposé englobe certaines lois ne possédant aucun moment , or contrairement à certaines idées reçues nous ne voyons aucune raison de répudier les lois sans moyenne (ou sans variance), lois dont l'usage peut apparaître judicieux dans certaines questions.

Il ne faudrait toutefois pas en conclure hâtivement que l'on peut utiliser le test F sans précautions, il faut en particulier observer que le lemme 6 du §1 marque une sérieuse limitation pour qui tient à l'indépendance ; dans le même ordre d'idée signalons le fait,(facile à établir), qu'un vecteur aléatoire du second ordre, sphériquement distribué autour de m, ($m \in R^n$), admet obligatoirement m comme moyenne et possède une matrice de covariance du type $\sigma^2 I_n$, excluant ainsi des corrélations entre ses coordonnées. En revanche, on notera que le fait pour un vecteur aléatoire d'être isotropiquement distribué n'interdit nullement l'existence de corrélations entre ses coordonnées.

Il resterait, bien sûr, à exhiber des modèles statistiques pertinents où l'isotropie de la distribution ne se réduirait pas à une invariance par rotation : l'exemple 3 du §1 n'a pas cette prétention.

Situation de ce travail par rapport à la littérature existante

Il convient tout d'abord de dire que c'est en 1973 que nous avons eu l'idée des principaux résultats contenus dans ce texte, en particulier de la proposition du §2.

Bien qu'ayant évoqué le sujet en petit comité, nous avons longtemps cru que "tout ceci étant assez simple, ce devait sûrement être des choses bien connues cachées quelque part dans la littérature"...

Après une communication privée, Ducimetière (I.N.S.E.R.M.) m'a fait remarquer, deux ans plus tard, que des préocupations voisines des notres existaient dans l'ouvrage [4] de Dempster. Il est parfaitement exact que le théorème 12.2.3 de cet auteur qui n'est qu'une variante de notre lemme 5, dans le cas où le vecteur est sphériquement distribué, est tout à fait au coeur du sujet. Ceci dit il nous a semblé que Dempster n'allait jamais au-delà des lois sphériquement distribuées et n'avait pas vu, non plus, le parti que l'on peut tirer des choses pour le test F du modèle linéaire. Sur ces points, nous croyons que notre proposition (§2) va nettement au-delà de ce qui apparaît dans son ouvrage.

Les réactions récentes aux exposés que nous avons fait sur cette question en différents lieux, en particulier à Aix-en-Provence (Mai 1976) aux journées de l'Association des Statisticiens Universitaires et à Grenoble (Septembre 1976) au Congrès Européen des Statisticiens, a, en tout cas, très largement fait reculer notre idée première selon laquelle tout ceci était sûrement "bien connu", même si certains aspects l'étaient en effet : ainsi, si Dempster, se limitant sur ces points à donner des arguments heuristiques, ne démontre pas rigoureusement, dans son ouvrage ce qui correspond essentiellement à nos lemmes 1 et 2, en revanche il renvoie à un excellent article de James [5] dans lequel se trouve énoncé un théorème général éclairant tout à fait la situation. Notre lemme 6 (ou une forme voisine) est assurément une propriété "bien connue", puisque déjà Rao [6] énonce (au début de son chapitre 3) un résultat de cette nature : c'est en particulier pour épargner au lecteur statisticien un effort bibliographique appréciable que nous avons opté pour un exposé autonome s'efforçant d'énoncer clairement tout ce qui est nécessaire en donnant pratiquement toutes les démonstrations.

3. DEMONSTRATIONS DES ENONCES DES PARAGRAPHES 1 ET 2

<u>Démonstration du Lemme 1</u>

Nous établirons successivement l'existence et l'univité de σ_{n-1}.

a) <u>existence de σ_{n-1}</u>

On désigne par λ_n la mesure de Lebesgue de R^n, par

$V_n = \lambda_n [\{x \ ; \ x \in R^n \text{ et } ||x|| \leqslant 1 \}]$ le volume de la boule

unité de R^n, par T l'application de R^n-{0} dans R^+ définie par

(3.1) $$T(x) = x/||x|| \quad , \text{ pour tout } x \in R^n \text{ , } x \neq 0$$

(T est ainsi l'application donnant l'angle polaire de x), par σ_{n-1} la mesure sur S_{n-1} définie par

(3.2) $$\sigma_{n-1}(B) = \frac{1}{V_n} \lambda_n(T^{-1}(B)) \text{ , pour tout B borélien}$$

de S_{n-1}.

Il est clair que σ_{n-1} est une probabilité, l'invariance par rotation provient de ce que T commute avec le groupe O(n), en ce sens que

(3.3) $$A \circ T = T \circ A \quad \text{pour tout } A \in O(n) \text{ ;}$$

de sorte que , pour tout borélien B de S_{n-1} et tout $A \in O(n)$,

$$V_n \sigma_{n-1}(A^{-1}(B)) = \lambda_n(T^{-1}(A^{-1}(B)) = \lambda_n(A^{-1}(T^{-1}(B)))$$

$$= \lambda_n(T^{-1}(B)) \quad \text{(invariance de } \lambda_n \text{ par rotation),}$$

$$= V_n \sigma_{n-1}(B) \quad .$$

b) <u>unicité de σ_{n-1}</u>

Elle s'établit simplement en faisant appel à la théorie de la mesure de Haar, sujet sur lequel on peut, par exemple, consulter [7] ou [8] ; pour l'usage que nous en avons, nous résumerons en un "Lemme" les résultats que nous emprunterons à cette théorie :

<u>Lemme</u> :

Le groupe O(n) est un groupe topologique compact, sur lequel il existe une unique mesure de probabilité ν invariante par les translations à gauche, c'est-à-dire par les applications $\Omega \to A\Omega$, pour tous A, $\Omega \in O(n)$; ν est appelée la mesure de Haar de O(n).

Ce lemme acquis, passons à l'unicité de σ_{n-1}. Pour toute fonction $f \in \mathcal{C}(S_{n-1})$, ensemble des fonctions continues sur S_{n-1} à valeurs réelles, et pour tout $\Omega \in O(R^n)$, on introduira les applications f_x(resp. f_Ω) définies sur $O(n)$ (resp. S_{n-1}), à valeurs dans R, par

(3.4) $$f_x(\Omega) = f_\Omega(x) = f(\Omega^{-1} x) \quad , \text{ pour tout } x \in S_{n-1} \text{ et}$$
tout $\Omega \in O(n)$.

Si μ est, alors, une probabilité sur S_{n-1}, invariante par rotation, et que $\mu(f)$ désigne l'intégrale de f par rapport à μ, on observe que, pour tout $f \in \mathcal{C}(S_{n-1})$ et tout $\Omega \in O(n)$:

(3.5) $$\mu(f) = \mu(f_\Omega) = \nu\,[\mu(f_\Omega)] = \mu\,[\nu(f_x)\,] \quad ,$$

la première égalité provient de l'invariance par rotation de μ, la seconde s'obtient en intégrant la fonction $\Omega \to \mu(f_\Omega)$ sur S_{n-1} par rapport à ν, la dernière en appliquant le théorème de Fubini à $\mu \otimes \nu$.

On achève en constatant que $\nu(f_x)$ est indépendante de $x \in S_{n-1}$, $O(n)$ opérant transitivement sur S_{n-1} (1); on définit donc une probabilité $\hat{\mu}$ sur S_{n-1} par la relation :

(3.6) $$\hat{\mu}(f) = \nu(f_x) \quad , \text{ pour tout } f \in \mathcal{C}(S_{n-1}), \text{ tout } x \in S_{n-1}.$$

Les relations (3.5) et (3.6) donne alors $\mu = \hat{\mu}$, ce qui établit l'unicité souhaitée.

---

(1) la transitivité signifie que pour tous $x, y \in S_{n-1}$, il existe $A \in O(n)$ tel que $x = Ay$, donc $\nu(f_x) = \nu(f_{A^{-1}y})$ $= \nu(f_y)$, en utilisant l'invariance de ν par A.

Remarque :

On notera qu'en fait la démonstration d'unicité résoud aussi le problème de l'existence, car la mesure $\hat{\mu}$ définie par (3.6) répond clairement à la question. Il nous a toutefois paru intéressant d'établir l'existence de σ_{n-1} par une construction géométrique élémentaire.

Démonstration du Lemme 2

a) (jj) $\Rightarrow$ (j)

Soit Y satisfaisant (jj), on désigne par ν la loi de probabilité (sur R^+) de $||Y||$, soit f une fonction continue bornée sur R^n, $A \in O(n)$; on obtient par un passage en coordonnées polaires :

$$\int_{R^n} f(Ay)\; P_Y(dy) = \int_{R^+} \nu(d\rho) \int_{S_{n-1}} f(A\rho\theta)\sigma_{n-1}(d\theta)$$

$$= \int_{R^+} \nu(d\rho) \int_{S_{n-1}} f(\rho\theta)\; \sigma_{n-1}(d\theta) = \int_{R^n} f(y)\; P_Y(dy),$$

en observant que $A\rho\theta = \rho A\theta$ puis en utilisant l'invariance de σ_{n-1}, d'où le résultat.

b) (j) $\Rightarrow$ (jj)

Soit Y satisfaisant (j), C un borélien de R^+ tel que $P[\,||Y|| \in C\,] > 0$; on définit sur S_{n-1} une mesure de probabilité μ_C par

(3.7) $$\mu_C(.) = P[\,||Y|| \in C,\; Y/||Y|| \in .\,]/P[\,||Y|| \in C],$$

μ_C est invariante par O(n) ; si en effet $A \in O(n)$ et si B est un borélien de S_{n-1} :

$$\mu_C(A^{-1}(B)) = P[\,||Y|| \in C,\ Y/||Y|| \in A^{-1}(B)]\,/\,P[\,||Y|| \in C]$$
$$= P[\,||AY|| \in C,\ AY/||AY|| \in B]\,/\,P[\,||AY|| \in C]$$
$$= \mu_C(B)\quad ,$$

en utilisant successivement le fait que A conserve la norme et l'invariance de la loi de Y.

D'après le lemme 1, $\mu_C = \sigma_{n-1}$; donc, pour tout C, borélien de R^+, et tout B, borélien de S_{n-1}, on obtient d'après (3.7) :

(3.8) $$P[\,||Y|| \in C,\ Y/||Y|| \in B\,] = P[\,||Y|| \in C\,]\ \sigma_{n-1}(B)\ ;$$

faisant $C = R^+$, on en déduit en premier lieu que σ_{n-1} est la loi de $Y/||Y||$, d'où le résultat en revenant à (3.8).

Démonstration du lemme 3

On désigne par I l'identité de R^n, par d le rang de Π, on note O(d) le groupe orthogonal de Im Π ;

Propriété (k)

Supposons que Y soit sphériquement distribué dans R^n. Pour tout $a \in O(d)$ on associe clairement un élément $A \in O(n)$ par la formule :

(3.9) $$A(y) = a(\Pi y) + (I-\Pi)\,y \quad , \text{ pour tout } y \in R^n\ ;$$

or $\Pi A = a\Pi$, de sorte que, pour tout borélien $B \subset \mathrm{Im}\,\Pi$, on obtient successivement :

$$P[\,a\Pi Y \in B\,] = P[\,\Pi AY \in B\,] = P[\,AY \in \Pi^{-1}(B)]$$
$$= P[\,Y \in \Pi^{-1}(B)] \text{ (par invariance de } P_Y)$$
$$= P[\,\Pi Y \in B\,] \text{ , d'où le résultat .}$$

Propriété (kk)

Supposons que Y soit isotropiquement distribué dans R^n, ceci entraîne que $Y/||Y||$ est sphériquement distribué, donc, avec (k), $\Pi(Y/||Y||) = \Pi Y/||Y||$ est sphériquement distribué, donc isotropiquement distribué dans Im Π, si du moins l'on sait établir que $\Pi Y \neq 0$ p.s ; il ne reste plus qu'à observer que l'angle polaire de $\Pi Y/||\Pi Y||$ est aussi celui de $\Pi Y/||Y||$ pour obtenir (kk).

Reste à établir que $\Pi Y \neq 0$ p.s. Pour ce faire on constate que $P[\Pi Y = 0] = P[Y \in \text{Ker } \Pi] = P[Y/||Y|| \in \text{Ker } \Pi]$
$$= \sigma_{n-1}(S_{n-1} \cap \text{Ker } \Pi)$$

or, utilisant l'explicitation de σ_{n-1} fournie par la démonstration d'existence du lemme 1, il vient

$$P[\Pi Y = 0] = \frac{1}{V_n} \lambda_n \{x \; ; \; x \in R^n, ||x|| \leq 1 \text{ et } x \in \text{Ker } \Pi\}$$

mais l'hypothèse rg $\Pi > 1$ assure que Ker Π, s.e.v de dimension strictement inférieur à n, est λ_n-négligeable, de sorte que $P[\Pi Y = 0] = 0$.

Démonstration du lemme 4

Nous démontrerons la propriété pour une composante particulière de Y, par exemple Y_1. Supposons donc que Y soit isotropiquement distribué dans R^n $(n \geq 2)$, en vertu du lemme 3, Y_1 est isotropiquement distribué dans R, de sorte que

(3.10) $$P[Y_1/||Y_1|| = \pm 1] = 1/2$$

si bien que 0 est une médiane de la loi de Y_1. L'unicité de cette médiane est moins facile à obtenir, nous raisonnerons par l'absurde.

Soit $a \neq 0$, une autre médiane de la loi P_{Y_1}, (p. ex. $a > 0$ pour fixer les idées), nous avons simultanément $P[Y_1 > 0] = 1/2$ et $P[Y_1 \geqslant a] \geqslant 1/2$, si bien que $0 = P[0 < Y_1 < a] = P[(Y_1, Y_2) \in C]$, en désignant par C le pavé ouvert $C =]0,a[\times R$, $(C \subset R^2)$.

Pour tout $k > 0$, soit $B(k)$ la boule fermée de rayon k dans R^2 ; on observe que $B(k)$ est recouverte par les ouverts $\{A(C) \; ; \; A \in O(2)\}$: un nombre fini, les $(A_i(C) \; ; \; 1 \leqslant i \leqslant N)$, suffit à recouvrir $B(k)$. Par conséquent

$$P[(Y_1, Y_2) \in B(k)] \leqslant P[(Y_1, Y_2) \in \bigcup_{i=1}^{N} A_i(C)]$$

(3.11)
$$\leqslant \sum_{i=1}^{N} P[(Y_1, Y_2) \in A_i(C)]$$

Mais, utilisant l'invariance de la loi de (Y_1, Y_2), on observe que, pour tout $i < N$,

$$P[(Y_1, Y_2) \in A_i(C)] = P[(Y_1, Y_2) \in C] = 0 ,$$

revenant à (3.11), on en déduit que

$P[(Y_1, Y_2) \in B(k)] = 0$, pour tout $k > 0$, ce qui est clairement impossible.

Démonstration du lemme 5

Soit Y un vecteur isotropiquement distribué dans R^n, au vu de la définition 2, la loi de probabilité de l'angle polaire $\Theta = Y/||Y||$ est σ_{n-1}. Or

(3.12)
$$\frac{||\Pi_1 Y||^2/d_1}{||\Pi_2 Y||^2/d_2} = \frac{||\Pi_1 \Theta||^2/d_1}{||\Pi_2 \Theta||^2/d_2}$$
, de sorte que le

lemme sera établi dès lors que l'on aura montré que la loi de

Fisher-Snédécor $F(d_1, d_2)$ est l'image de σ_{n-1} par l'application

(3.13) $$\theta \to \frac{\|\Pi_1 \theta\|^2/d_1}{\|\Pi_2 \theta\|^2/d_2}$$

Pour établir ceci, il suffit de considérer le cas d'un vecteur aléatoire $Y = (Y_1, \dots, Y_{d_1}, Y_{d_1+1}, \dots, Y_{d_1+d_2}, Y_{d_1+d_2+1}, \dots Y_n)$ où les n v.a. Y_k $(n \geqslant d_1 + d_2)$ suivent une même loi de Gauss centrée réduite, c'est un résultat classique (souvent pris comme définition) que de vérifier que

$$F(d_1, d_2) \text{ est la loi de } \frac{(Y_1^2 + \dots + Y_{d_1}^2)/d_1}{(Y^2_{d_1+1} + \dots + Y^2_{d_1+d_2})/d_2} = \frac{\|\Pi_1 Y\|^2/d_1}{\|\Pi_2 Y\|^2/d_2},$$

en désignant par Π_1 et Π_2 les projecteurs orthogonaux de R^n associant à $y, (y \in R^n)$, les éléments :

$$\Pi_1\ y = (y_1, \dots, y_{d_1},\ 0\ , \dots,\ 0\ , 0, \dots, 0)$$
$$\Pi_2\ y = (0\ , \dots,\ 0\ ,\ y_{d_1+1}, \dots, y_{d_1+d_2}, 0, \dots, 0)$$

utilisant le fait que le vecteur gaussien Y satisfait à la condition (j) du lemme 2, on note que son angle polaire $\Theta = Y/\|Y\|$ suit σ_{n-1} par l'application définie en (3.13).

<u>Démonstration du lemme 6</u>

Supposons que la loi P_Y de $Y = (Y_k\ ;\ 1 \leqslant k \leqslant n)$ soit invariante par $O(n)$; ceci entraîne clairement que les Y_k possèdent, pour tout k, une même loi $P_{Y_k} = p$; d'après le lemme 3, la loi du couple (Y_1, Y_2) est invariante par $O(2)$ (groupe des rotations planes), d'où résulte que pour tout $\alpha \in (0, 2\Pi)$, p est aussi la loi de $Y_1 \cos\alpha + Y_2 \sin\alpha$, Si nous désignons par φ la transformée de Fourier de p, $\varphi(t) = \int_R e^{ity} p(dy)$, on obtient, utilisant l'indépendance de Y_1 et Y_2 :

(3.14) $$\varphi(t) = \varphi(t \cos\alpha)\ \varphi(t \sin\alpha)\ ,\ \text{pour tout } t \in R,$$

et tout $\alpha \in (0,2\Pi)$; changeant α en $\alpha+\Pi$, il vient aussi

$$\varphi(t) = \varphi(-t \cos\alpha)\ \varphi(-t \sin\alpha)\ ,\ \text{si bien que}$$

(3.15) $$\varphi(t) = \varphi(-t),\ \text{pour tout } t \in R\ ;$$

comme, identiquement, $\overline{\varphi(t)} = \varphi(-t)$, on constate que φ est une fonction paire à valeurs réelles ; mais φ est continue, il existe donc une fontion Ψ continue de R^+ dans lui-même telle que :

(3.16) $$\varphi(t) = \Psi(t^2)\ ,\ \text{pour tout } t \in R\ ;$$

grâce à (3.14), Ψ satisfait, identiquement, à

$$\Psi(t^2) = \Psi(t^2\cos^2\alpha)\Psi(t^2\sin^2\alpha)\ ,\ \text{soit aussi à}$$

(3.17) $$\Psi(a+b) = \Psi(a)\ \Psi(b)\ ,\ \text{pour tous } a,b \in R^+\ ;$$

par conséquent, pour une constante K convenable,

$$\Psi(t) = \exp(Kt)\ ,\ \text{de sorte que}$$
$$\varphi(t) = \exp(Kt^2)\ ,\ \text{pour tout } t \in R.$$

Comme il est nécessaire que $\varphi(t)$ reste inférieur à 1, il existe un $\sigma^2 \geqslant 0$ tel que $K = -\sigma^2$; la loi commune p des Y_k est bien une gaussienne centrée.

Démonstration de la proposition du §2

Nanti des résultats du §1, la démonstration est vite faite :

Soit Y un vecteur aléatoire satisfaisant les hypothèses de la proposition, on remarque, que, quand $m \in L_0$, la statistique F définie en (2.2) s'écrit encore

$$F = \frac{||(\Pi-\Pi_0)(Y-m)||^2/(\ell-\ell_0)}{||(I_n-\Pi)(Y-m)||^2/(n-\ell)} \quad ;$$

les hypothèses du lemme 5 (§1) sont satisfaites par le vecteur aléatoire Y-m et par les projecteurs orthogonaux $\Pi - \Pi_0$ et $I_n - \Pi$ dont les rangs respectifs sont $\ell-\ell_0$ et $n-\ell$ d'où le résultat.

x

x x

REFERENCES BIBLIOGRAPHIQUES

[1] JOHNSON and KOTZ "Distributions in statistics : continuous multivariate distributions", Wiley, 1972.

[2] E.L. LEHMANN "Testing Statistical Hypotheses", Wiley, 1965.

[3] G.A.F. SEBER "The Linear Hypotheses : a general theory", Griffin, 1966.

[4] A.P. DEMPSTER "Elements of continuous multivariate analysis", Addison-Wesley, 1969.

[5] A.T. JAMES "Normal multivariate analysis and the orthogonal group", Ann. of Math. Stat., vol.25, 1964, p. 40 à 75.

[6] C.R. RAO "Linear Statistical Inference and its Applications", Wiley, 1965.

[7] LOOMIS "An introduction to abstract harmonic analysis", Van Nostrand, 1953.

[8] A. WEIL "L'intégration dans les groupes topologiques et ses applications", Hermann, 1953.

Jean-Louis PHILOCHE
Le Houssais
35740 PACE.