



HAL
open science

Extrémalité et stabilité des composantes principales (Généralisation de l'étude euclidienne)

C. Deniau, Georges Oppenheim

► **To cite this version:**

C. Deniau, Georges Oppenheim. Extrémalité et stabilité des composantes principales (Généralisation de l'étude euclidienne). Annales de l'ISUP, 1976, XXI (1-2), pp.27-41. hal-04082279

HAL Id: hal-04082279

<https://hal.sorbonne-universite.fr/hal-04082279>

Submitted on 26 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

EXTREMALITE ET STABILITE DES COMPOSANTES PRINCIPALES (GENERALISATION DE L'ETUDE EUCLIDIENNE)

C. DENIAU

G. OPPENHEIM

SUMMARY

Given a real matrix $X \in \{n \times p\}$, optimality properties of the principal axes and their associated components are demonstrated. Each property is tied to a criterion which depends on one or two scalar products (the choice of which can, in some cases, be left to the research worker). Some characteristics of these properties have been given by several authors : RAO (1964), DARROCH (1965), OKAMOTO (1968) ; they are founded on the maximisation of adjustment qualities, the minimisation of adjustment residuals and the maximisation of volumes (generalised variance). The author of the present paper have grouped and generalized preceding results in an Euclidian frame, a probabilistic context being here unnecessary.

All the results are presented in the form of necessary and sufficient conditions (C.N.S.) and use the properties of the generalized Gramm matrices $\Gamma_n = X \Psi^t X \Phi$, $\Gamma_p = {}^t X \Psi X \Phi$ associated to X and to two symmetrical definite positive matrices. Several applications are submitted in part 4. In part 1, notations and definitions have been gathered and, in part 2, the necessary and sufficient conditions bearing on extremality properties of quadratic functions and determinants. Part 3 uses the results demonstrated in part 2 for a study of stability and extremality of the principal components.

INTRODUCTION

Le but de cet article est de présenter dans un cadre algébrique général quelques propriétés d'optimalité (extrémalité et stabilité) des composantes principales d'une matrice $X \in \{n \times p\}$.

Différentes caractérisations de ces propriétés ont été données dans plusieurs articles, par exemple : ANDERSON [1], RAO [11] caractérisent les composantes principales de $X \in \{n \times p\}$ par des propriétés d'optimalité portant sur l'ensemble des composantes de cette matrice ; DARROCH [4] et OKAMOTO [10] les caractérisent par des propriétés d'optimalité dans tout l'espace $\{p \times 1\}$.

Nous commençons par généraliser ces propriétés en considérant :

- a) toute matrice $X \in \{n \times p\}$ comme une application linéaire de $\{p \times 1\}$ dans $\{n \times 1\}$
- b) deux matrices symétriques définies positives $\phi \in \{n \times n\}$ et $\Psi \in \{p \times p\}$ munissant $\{n \times 1\}$ et $\{p \times 1\}$ chacune d'une structure euclidienne. (Lorsque $\phi = I_n$ et $\Psi = I_p$ nous retrouvons les résultats classiques).

Après des notations et définitions (1') des conditions nécessaires et suffisantes (2') permettent de démontrer les propriétés d'optimalité (3'). Nous terminons par des remarques et des applications. Nous avons délibérément choisi de nous placer dans un contexte algébrique non probabiliste ; en effet, d'une part le passage aux probabilités ne pose pas de problème particulier dès que l'on suppose que les v.a. considérées sont de carré sommable, et d'autre part nous n'abordons pas les problèmes de test et d'estimation.

1 – NOTATIONS ET PRELIMINAIRES

- $\{n \times p\}$ espace vectoriel des matrices réelles à n lignes, p colonnes.
- $\{n \times p\}_r$ ensemble des matrices réelles à n lignes et p colonnes, de rang r ($r \leq \inf(n, p)$)
- $A \in \{p \times p\}$ on note $\text{diag } A$ la matrice diagonale dont les éléments sont les éléments diagonaux de A .
- $X \in \{n \times p\}$ on note $r(X)$ le rang de X .
- $B \in \{p \times p\}$ symétrique positive ; on notera toujours $\lambda_1(B) \geq \dots \geq \lambda_p(B) \geq 0$ les valeurs propres de B . Quel que soit $k \in]p]$:
 $\lambda_1(B) ; \dots ; \lambda_k(B) ; (\text{resp } \lambda_{p-k+1}(B), \dots, \lambda_p(B))$
sont les k premières (resp k dernières) valeurs propres de B .
- soit (u_1, \dots, u_k) une famille libre de vecteurs de $\{n \times 1\}$; on la notera U_k et on l'identifiera à la matrice $\{n \times k\}$ dont les k colonnes sont respectivement

ment u_1, \dots, u_k . On notera $[U_k]$ ou $[u_1, \dots, u_k]$ le sous-espace vectoriel de $\{n \times 1\}$ engendré par u_1, \dots, u_k .

- $\mathcal{E}_n = \{ \{n \times k\} ; k \in \mathbb{N} \}$

- quel que soit $k \in \mathbb{N}^* :]k[= \{1, \dots, k\}$

- $S_n = \{x \in \{n \times 1\}; \|x\|_\phi = 1\}$; $S_p = \{y \in \{p \times 1\} ; \|y\|_\phi = 1\}$

- $I_k \in \{k \times k\}$ est la matrice identité.

Définition (1.1.)

- a) soit $A \in \{n \times n\}$; on dit que A est ϕ -symétrique si et seulement si $B = \phi A$ est symétrique.
- b) soit $P \in \{n \times n\}$; on dit que P est ϕ -orthogonale si et seulement si ${}^t P \phi P = I_n$.

Lemme (1.1)

Soit $A \in \{n \times n\}$ ϕ -symétrique positive. Alors il existe A unique et $V \in \{n \times n\}$ telles que :

$$AV = V\Lambda ; {}^t V \phi V = I_n ; \Lambda = \text{diag } \Lambda. \tag{1.1}$$

Les éléments diagonaux de $\Lambda : \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ sont les valeurs propres de A, les colonnes v_1, \dots, v_n de V sont des vecteurs propres ϕ -orthogonaux de A associés respectivement aux valeurs propres $\lambda_1, \dots, \lambda_n$ de A. La matrice V n'est pas unique. Si V est une solution particulière de (1.1), toute solution (1) de (1.1) est la forme VR avec :

$$R \in \{n \times n\}, \quad R = \begin{bmatrix} [\rho_1] & & & 0 \\ & \ddots & & \\ 0 & & [\rho_s] & \\ & & & \ddots \end{bmatrix} ; {}^t R R = I_n, \tag{1.2}$$

est le nombre des valeurs propres de A distinctes. Pour tout $h \in]s[: n_h$ est l'ordre de multiplicité de λ_h et $\rho_h \in \{n_h \times n_h\}$.

(1) (l'idée d'introduire cette matrice vient d'OKAMOTO [10])

30.

Définition (1.2)

Soit $X \in \{n \times p\}_r$, on appelle matrice de Gramm-généralisée associée à X la matrice ϕ -symétrique $\Gamma_n = X \psi^t X \phi$.

A la matrice ${}^t X \in \{p \times n\}$ on associe la matrice de Gramm-généralisée ψ -symétrique: $\Gamma_p = {}^t X \phi X \psi$.

Lemme 1.2.

Soit $\phi \in \{n \times n\}$ (resp. $\psi \in \{p \times p\}$) symétrique définie positive, il existe $\phi_1 \in \{n \times n\}$ (resp. $\psi_1 \in \{p \times p\}$) régulière telle que :

$$\phi = {}^t \phi_1 \phi_1 \quad (\text{resp. } \psi = {}^t \psi_1 \psi_1).$$

Lemme (1.3)

Soient $X \in \{n \times p\}_r$, Γ_p et Γ_n .

Les matrices Γ_p et Γ_n sont toutes deux de rang r et ont les mêmes valeurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ non nulles. La valeur propre nulle a pour ordre de multiplicité $(p - r)$ pour Γ_p , $(n - r)$ pour Γ_n .

Pour tout $i \in]r, p]$ notons $v_i \in \{p \times 1\}$ (resp. $w_i \in \{n \times 1\}$) un vecteur propre ϕ -unitaire (resp. ψ -unitaire) de Γ_n (resp. Γ_p) associé à λ_i , alors ;

$$v_i = (\lambda_i)^{-1/2} X \psi w_i \quad ; \quad w_i = (\lambda_i)^{-1/2} {}^t X \phi v_i. \quad (1.3)$$

Lemme (1.4)

L'application qui a $(X, Y) \in \{n \times p\}^2$ associe $\text{Tr} ({}^t X \phi Y \psi)$

définit un produit scalaire sur $\{n \times p\}$

On notera $\|X\|$ la norme de X associée à ce produit scalaire. (Notons que si $\phi = I_n$ et $\psi = I_p$ on retrouve le produit scalaire usuel).

Définition (1.3)

a) Soient $X \in \{n \times p\}$ et $u \in S_n$. On appelle composante de X associée à u :

$$C = {}^t X \phi u \in \{p \times 1\}.$$

b) On appelle k -ième composante principale de X ($k \leq r(X)$), la composante de X associée à v_k ($v_k \in S_n$) vecteur propre de Γ_n associé de λ_k .

Remarque 1.1.

Les lemmes (1.1), (1.3), se déduisent simplement du cas où $\phi = I_n$ (RAO [12]):

- i) Il existe $\phi_1 \in \{n \times n\}$ régulière telle que ${}^t\phi_1\phi_1 = \phi$.
 Tout $A \in \{n \times n\}$ ϕ -symétrique est de la forme $A = \phi^{-1} B$ avec B symétrique
- ii) quel que soit $i \in]n[$: $\lambda_i(A) = \lambda_i({}^t\phi_1^{-1}B\phi_1^{-1})$;
 si u_i est vecteur propre de ${}^t\phi_1^{-1}B\phi_1^{-1}$ associé à λ_i , $V_i = \phi_1^{-1}u_i$ est vecteur propre de A associé à la même valeur propre.
- iii) si $(u_i \mid i \in]n[)$ est une famille de vecteurs propres orthonormés de ${}^t\phi_1^{-1}B\phi_1^{-1}$, $(v_i \mid i \in]n[)$ est une famille de vecteurs propres ϕ -orthonormés de A .

2 - CONDITIONS NECESSAIRES ET SUFFISANTES

C.N.S.1. Soit $A \in \{n \times n\}$ positive ϕ -symétrique et $B = \phi A$.

1. Quelque que soit $u \in S_n$: ${}^t u B u \leq \lambda_1(A)$ (2.1)

* Une condition nécessaire et suffisante pour que ${}^t u B u = \lambda_1(A)$ est que $u = v_1$, vecteur propre de A associé de $\lambda_1(A)$.

2. Pour tout $k \in]p-1[$, soit (v_1, \dots, v_k) une famille de vecteurs propres ϕ -orthonormés de A .

Quel que soit $u \in S_n \cap [v_1, v_2, \dots, v_k]^\perp$: ${}^t u B u \leq \lambda_{k+1}(A)$.

* Une condition nécessaire et suffisante pour que ${}^t u B u = \lambda_{k+1}(A)$ est que $u = v_{k+1}$, vecteur propre de A associé à $\lambda_{k+1}(A)$.

3. Soit $U_k \in \{n \times k\}$ ($k \leq n$) telle que ${}^t U_k \phi U_k = I_k$.

$$\forall i \in]k[: \lambda_i({}^t U_k B U_k) \leq \lambda_i(A). \quad (2.2)$$

* Une condition nécessaire et suffisante pour que :

$$\forall i \in]k[: \lambda_i({}^t U_k B U_k) = \lambda_i(A)$$

est que $U_k = V R_{(k)} P$, où $V \in \{n \times n\}$ est une matrice de vecteurs propres ϕ -orthonormés de A , $R_{(k)} \in \{n \times k\}$ est constituée par les k premières colonnes de R (lemme 1.1) et $P \in \{k \times k\}$ telle que ${}^t P P = I_k$.

32.

4. Soient $k \in]p]$ et $U_k = (u_1, \dots, u_k)$ une famille ϕ -orthogonale de S_n :

$$\sum_{1 \leq i \leq k} {}^t u_i B u_i \leq \sum_{1 \leq i \leq k} \lambda_i(A) \quad (2.3)$$

* Une condition nécessaire et suffisante pour que l'on ait égalité en (2.3) est que $U_k = VR_{(k)}P$.

Preuve : La transformation $u' = \phi_1 u$ (Remarque 1.1) permet de se ramener à des conditions classiques (OKAMOTO [10], RAO [12]) :

$$S_n = \{u \mid {}^t u \phi u = 1\} = \{u' \mid {}^t u' u' = 1\} \text{ et } {}^t u B u = {}^t u' {}^t \phi_1^{-1} B \phi_1^{-1} u'$$

Or pour u' tel que ${}^t u' u' = 1$ ${}^t u' {}^t \phi_1^{-1} B \phi_1^{-1} u' \leq \lambda_1 ({}^t \phi_1^{-1} B \phi_1^{-1}) = \lambda_1(A)$.

Cette remarque achève la démonstration.

C.N.S.2 Soit $A \in \{n \times n\}$ positive ϕ -symétrique et $B = \phi A$;

Soit $U_k \in \{n \times k\}$ ($k \leq n$) telle que $\text{diag } {}^t U_k \phi U_k = I_k$, alors :

i) $\det {}^t U_k B U_k \leq \prod_{i \leq i \leq k} \lambda_i(A)$

ii) si $r(B) \geq k$, une condition nécessaire et suffisante pour que l'on ait égalité en (2.4) est que $U_k = VR_{(k)}P$.

Preuve :

1) Soient $\delta_1 \geq \dots \geq \delta_k$ les valeurs propres de ${}^t U_k \phi U_k$, Δ_k la matrice diagonale et $H = (h_i \mid i \in]k])$ une famille de vecteurs propres associés.

On choisit H telle que ${}^t H H = I_k$.

– Si $r(U_k) < k$ la démonstration de i) est achevée.

– Supposons $r(U_k) = k$; posons $G = H A^{-1/2} \in \{k \times k\}$, $Q = U_k G \in \{n \times k\}$

$${}^t Q \phi Q = \Delta_k^{-1/2} {}^t H {}^t U_k \phi U_k H \Delta_k^{-1/2} = I_k$$

$$\text{Tr}({}^t U_k \phi U_k) = k, \text{ donc } : \left(\prod_{1 \leq i \leq k} \delta_i \right)^{1/k} \leq (1/k) \text{Tr} \Delta_k \text{ et}$$

$$\det {}^t G \det G = \left(\prod_{1 \leq i \leq k} \delta_i \right)^{-1} \geq 1$$

$$\det({}^tQBQ) = \det({}^tG{}^tU_kBU_kG) = \det(G{}^tG) \det({}^tU_kBU_k) \geq \det({}^tU_kBU_k) \quad (2.5)$$

donc d'après (C.N.S. 1) : $\det({}^tQBQ) = \prod_{1 \leq i \leq k} \lambda_i({}^tQBQ) \leq \prod_{1 \leq i \leq k} \lambda_i(A)$ (A)

$$\det({}^tU_kBU_k) \leq \det({}^tQBQ) \leq \prod_{1 \leq i \leq k} \lambda_i(A). \quad (2.6)$$

2. Condition suffisante

Si $U_k = VR_{(k)} P$, d'après (C.N.S.1) :

$$\begin{aligned} \det({}^tU_kBU_k) &= \det({}^tP{}^tR_{(k)}{}^tVBVR_{(k)}P) = \det({}^tR_{(k)}{}^tVBVR_{(k)}) \\ &= \prod_{1 \leq i \leq k} \lambda_i(A). \end{aligned}$$

Condition nécessaire

Si $\det({}^tU_kBU_k) = \prod_{1 \leq i \leq k} \lambda_i(A)$, d'après (2.3), (2.5) et (2.6) quel que soit $i \in]k]$: $\delta_i = 1$ et $\lambda_i({}^tQBQ) = \lambda_i(A)$. Donc tout vecteur de $\{k \times 1\}$ est vecteur propre de ${}^tU_k \phi U_k$ et on peut poser $H = I_k$. Alors $G = I_k$, $Q = U_k$ et d'après (C.N.S.1) $U_k = VR_{(k)}P$.

Le théorème suivant propose une généralisation du théorème démontré par J.N. DARROCH [4].

C.N.S.3

Soient $X \in \{n \times p\}_r$, $G \in \{n \times k\}_k$, $H \in \{p \times k\}_k$, $k \leq r$.

$$a) \quad \|X - G{}^tH\|^2 \geq \sum_{i=k+1}^r \lambda_i(\Gamma_n). \quad (2.7)$$

b) Une condition nécessaire et suffisante pour que l'on ait égalité en (2.7)

est que $G{}^tH = \sum_{i=1}^k v_i {}^t c_i$ où $(v_i | i \in]k])$ est une famille de vecteurs propres

ϕ -orthonormaux associés aux k premières valeurs propres de Γ_n et $(c_i | i \in]k])$ la famille des composantes principales correspondantes.

34.

Preuve

$$\begin{aligned} \|X - G^t H\|^2 &= \text{Tr} [(X - G^t H) \psi^t (X - G^t H) \phi] \\ &= \text{Tr} [\phi_1 (X - G^t H)^t \psi_1 \psi_1^t (X - G^t H) \phi_1] \\ &= \text{Tr} [(Z - J^t K)^t (Z - J^t K)] \end{aligned}$$

avec $Z = \phi_1 X^t \psi_1$, $J = \phi_1 G$, $K = \psi_1 H$.

(2.8)

Le théorème de DARROCH appliqué aux matrices Z, J, K établit que :

$$\text{Tr} [(Z - J^t K)^t (Z - J^t K)] \geq \sum_{i=k+1}^r \lambda_i(\Gamma_n)$$

(car les valeurs propres de Γ_n et de $Z^t Z$ sont les mêmes).

$$\text{L'égalité ayant lieu si et seulement si : } J^t K = \sum_{i=1}^k u_i^t f_i; ({}^t u_i u_i = 1) \quad (2.9)$$

avec u_i vecteur propre de $Z^t Z$ associé à $\lambda_i(\Gamma_n)$ et $f_i = {}^t Z u_i$.Or $u_i = \phi_1 v_i$ ($1 \leq i \leq k$) où v_i est un vecteur propre ϕ -unitaire de Γ_n associé à $\lambda_i(\Gamma_n)$ (Remarque 1.1)

En tenant compte de (2.8) et (2.9) on peut écrire :

$$\phi_1 G^t H^t \psi_1 = \sum_{i=1}^k \phi_1 v_i^t v_i \phi X^t \psi_1 = \sum_{i=1}^k \phi_1 v_i^t c_i^t \psi_1 ;$$

$$\phi_1 \text{ et } \psi_1 \text{ étant régulières : } G^t H = \sum_{i=1}^k v_i^t c_i$$

ce qui achève la démonstration.

3. EXTREMALITE ET STABILITE DES COMPOSANTES PRINCIPALES

3.1. Ajustements

(3.1.1) Définition (3.1) : fonction critère θ

Soient $X = (x_1, \dots, x_p) \in \{n \times p\}$ et $\theta_X : \mathcal{E}_n \rightarrow \mathbb{R}_+$ définie par :

$$\forall [U_k] \in \mathcal{E}_n : \theta_X([U_k]) = \sum_{1 \leq i \leq p} \|P_{U_k}(x_i)\|^2 \psi \quad (3.1)$$

où P_{U_k} est l'opérateur de projection ϕ -orthogonale sur $[U_k]$.

Notations matricielles

On peut, sans nuire à la généralité de la définition (2.1) poser

${}^t U_k \phi U_k = I_k$ et écrire :

$$\theta_X([U_k]) = \sum_{1 \leq i \leq k} \| {}^t X \phi u_i \|^2$$

$$= \sum_{1 \leq i \leq k} {}^t u_i \phi X \psi {}^t X \phi u_i$$

$$\theta_X([U_k]) = \sum_{1 \leq i \leq k} {}^t u_i \phi \Gamma_n u_i \tag{3.2}$$

(3.1.2) Etude de maximum de la fonction θ

Théorème (3.1.1.)

a) Soit $u \in S_n$:

$$\theta_X([u]) \leq \lambda_1(\Gamma_n) \tag{3.3}$$

Une condition nécessaire et suffisante pour que l'on ait égalité en (3.3) est que $[u] = [v_1]$ droite vectorielle engendrée par v_1 (vecteur propre de Γ_n associé à $\lambda_1(\Gamma_n)$).

b) Soient $U_k = (v_1, \dots, v_k)$ une famille de vecteurs propres associée aux k premières valeurs propres de Γ_n et $u \in [U_k]^\perp \cap S_n$:

$$\theta_X([u]) \leq \lambda_{k+1}(\Gamma_n). \tag{3.4}$$

Une condition nécessaire et suffisante pour que l'on ait égalité en (3.4) est que $[u] = [v_{k+1}]$ droite vectorielle engendrée par v_{k+1} (vecteur propre de Γ_n associée à $\lambda_{k+1}(\Gamma_n)$).

Théorème (3.1.2.)

Soient $X \in \{n \times p\}_r$ et $k \in]r[$ (3.5)

Une condition nécessaire et suffisante pour que l'on ait égalité en (3.5) est que $U_k = VR_{(k)}P$ (où $V, R_{(k)}, P$ sont définies dans la C.N.S.1).

36.

P reuve des théorèmes (3.1) et (3.2) :

La matrice $\phi \Gamma_n \in \{n \times n\}_r$ est symétrique positive, on est dans les conditions d'application de la C.N.S.1. ce qui achève la démonstration.

3.2. Ajustement de matrices

Définition (3.2) : fonction critère ζ .

Soit $X \in \{n \times p\}$ $\zeta_X : \{n \times p\} \rightarrow \mathbb{R}_+$ définie par : $\forall Z \in \{n \times p\} : \zeta_X(Z) = \|X - Z\|^2$

Théorème (3.2)

Soient $X \in \{n \times p\}_r$ et $k \in]r]$.

$$Z_k \in \{n \times p\}_k : \|X - Z_k\|^2 \geq \sum_{i=k+1}^r \lambda_i(\Gamma_n) \quad (3.6)$$

Une condition nécessaire et suffisante pour que l'on ait égalité en (3.6) est que $Z_k = \sum_{i=1}^k v_i^t c_i$ (où v_i et c_i sont définis dans la C.N.S.3.)

Preuve

Quel que soit $Z_k \in \{n \times p\}_k$ il existe $G \in \{n \times k\}_k$, $H \in \{p \times k\}_k$ tels que : $Z_k = G^t H$ (la preuve est dans [5] ou [12]).

Ceci nous ramène dans les conditions de la (C.N.S.3) ce qui achève la démonstration.

3.3 Ajustement : critère de volume

Définition (3.3.1) :

soit $Z = (z_1, \dots, z_n) \in \{p \times n\}$ on appelle Ψ -volume engendré par z_1, \dots, z_n et on note $\text{vol } \psi(z_1, \dots, z_n)$:

$$\text{vol } \psi(z_1, \dots, z_n) = [\rho \det({}^t Z Z)]^{1/2} \quad \text{avec } \rho = \left(\frac{1}{p-1} \right)^n, p > 1$$

Définition (3.3.2) : Fonction critère η

Soient $X = (x_1, \dots, x_p) \in \{n \times p\}$ et $\eta_X : \mathcal{E}_n \longrightarrow \mathbb{R}_+$ définie par :

$$\forall [U_k] \in \mathcal{E}_n : \eta_X([U_k]) = \text{vol } \psi({}^t X \phi u_1, \dots, {}^t X \phi u_k)$$

Notations matricielles

Supposons que ${}^t U_k \phi U_k = I_k$

$$\begin{aligned} \eta_X([U_k]) &= (\rho \det ({}^t U_k \phi X \psi {}^t X \phi U_k))^{1/2} \\ &= \rho \det ({}^t U_k \phi \Gamma_n U_k)^{1/2} \end{aligned}$$

Théorème (3.3.)

Soient $k \leq r(\Gamma_n)$, $U_k \in \{n \times k\}$, $\text{diag} ({}^t U_k \phi U_k) = I_k$

$$\eta_X([U_k]) \leq \left[\prod_{1 \leq i \leq k} \lambda_i(\Gamma_n) \right]^{1/2} \quad (3.7)$$

Une condition nécessaire et suffisante pour que l'on ait égalité en (3.7) est que : $U_k = V R_{(k)} P$ (où $V, R_{(k)} P$ sont définies dans la C.N.S.1)

Preuve

La matrice $\phi \Gamma_n \in \{n \times n\}_r$ est symétrique: on se trouve dans les conditions d'application de la (C.N.S.2), ce qui achève la démonstration.

3.4 Stabilité

Soient $X \in \{n \times p\}_r$, $\mathcal{U} = (v_i | i \in]r]) \subset \{n \times 1\}$ une famille ϕ -orthogonale de vecteurs non nuls, $\mathcal{V} = (w_i | i \in]r]) \subset \{p \times 1\}$ une famille ψ -orthogonale de vecteurs non nuls.

On suppose que $[v_1, \dots, v_r]$ est invariant par Γ_n .

Une condition nécessaire et suffisante pour que :

$$\forall i \in]r] : {}^t X \phi [v_i] = [w_i]$$

38.

est que \mathcal{V} soit une famille orthogonale de vecteurs propres de Γ_n associés aux r premières valeurs propres.

Condition suffisante : c'est le lemme (1.2)

Condition nécessaire

a) $\Gamma_n v_i \neq 0$.

Si $\Gamma_n v_i = 0$ alors ${}^t v_i \phi X \psi {}^t X \phi v_i = 0$ et $\|w_i\|_\phi^2 = 0$ (impossible)

b) $\Gamma_n v_i = \sum_{1 \leq k \leq r} \alpha_k v_k$

$$\begin{aligned} \text{Calculons les } \alpha_k : {}^t v_j \phi \Gamma_n v_i &= \sum_{1 \leq k \leq r} {}^t v_j \phi \alpha_k v_k \\ &= \alpha_j ({}^t v_j \phi v_j). \end{aligned}$$

$$\text{Or } {}^t v_j \phi \Gamma_n v_i = ({}^t w_j \psi w_i) \delta_{ij},$$

donc seul $\alpha_i = \lambda_i$ est non nul et $\Gamma_n v_i = \lambda_i v_i$

ce qui achève la démonstration.

Remarque (3.4)

Une condition nécessaire et suffisante pour que $\forall i \in [r] \quad {}^t X \phi [v_i] = [w_i]$ et $X \psi [w_i] = [v_i]$ est que \mathcal{V} et \mathcal{W} soient 2 familles de vecteurs propres de Γ_n et Γ_p associées aux r premières valeurs propres.

4. Applications et remarques

4.1. Applications

4.1.1. Les résultats des parties précédentes sont établis pour des matrices ϕ et ψ symétriques définies positives qui peuvent être considérées comme associées à deux produits scalaires respectivement dans $\{n \times 1\}$ et $\{p \times 1\}$.

Certains choix particuliers de ces matrices sont d'usage courant.

Nous en présentons quelques uns pour lesquels Φ et ψ sont diagonales.

1 a. $\phi = I_n ; \psi = I_p$ (KENDALL et STUART [8]) ; SHETH [13])

1 b. $\psi = I_p ; \phi_{ii} = \left(\sum_{1 \leq j \leq p} x^2_{ij} \right)^{-1}$

Un cas intéressant est celui de l'analyse de données dichotomiques :

$\forall (i, j): x_{ij} \in \{0, 1\}$ (BURT [3] ; BENZECRI [2]). On a alors :

$$\Phi_{ii} = \left(\sum_{1 \leq j \leq p} x_{ij} \right)^{-1} = m_i^{-1} \quad \text{En notant } m_{ii'} = \sum_{1 \leq j \leq p} x_{ij} x_{i'j} :$$

$$\Gamma_n = ((m_{ii'}, m_i^{-1})) \quad 1 \leq i, i' \leq n.$$

Pour cette matrice $0 \leq \lambda_i(\Gamma_n) \leq 1$ pour tout $1 \leq i \leq n$.

Le premier vecteur propre est ${}^t v_1 = (m_1, m_1, \dots, m_n)$ avec $\lambda_1 = 1$;

la composante principale associée c_1 est telle que :

$${}^t c_1 = {}^t ({}^t X \phi v_1) = (m_1, \dots, m_j, \dots, m_p) \text{ avec } m_j = \sum_{1 \leq i \leq n} x_{ij} ;$$

2a. Si X est tel que $X \delta_p = 0$, se reporter à OKAMOTO [10]

2b. $\phi = I_n$; $\psi = \frac{1}{p} I_p$: Γ_n est la matrice de covariance associée à X .

(MATTER [9], GABRIEL [6])

Remarque : Si $x_{ij} \in \{0, 1\}$ et que l'on pose $\Phi_{ii} = m_i^{-1}$ et $\psi = I_p$, pour faire l'analyse de $Y = X \left(I_p - \frac{\delta_p {}^t \delta_p}{p} \right)$.

Avec les notations de 1.b ci-dessus, on a :

$$\Gamma'_{ii'} = \left(m_{ii'} - \frac{m_i m_{i'}}{p} \right) m_i^{-1} \quad 1 \leq i, i' \leq n$$

$0 \leq \lambda_i(\Gamma'_n) \leq 1$ pour $0 \leq i \leq n$; le vecteur propre ${}^t v_1$ de (1.b) est associé à

$\lambda(\Gamma'_n) = 0$; les autres vecteurs propres et valeurs propres de Γ'_n sont égaux à ceux de Γ_n .

4.1.2. Remarque

Les axes principaux engendrés par les vecteurs propres de Γ_n sont solutions de plusieurs problèmes d'extrémalité

(1) Cette référence nous a été communiquée par le Professeur M. REUHLIN

40.

$$\text{* qualité d'ajustement maximale : } \sup_{v \in S_n} \theta_X([v])$$

$$\text{* résidu d'ajustement minimum : } \inf_{v \in S_n} \theta'_X([v])$$

$$\text{avec } \theta'_X([X]) = \theta_X(\{n \times 1\}) - \theta_X([v])$$

$$\text{* variance généralisée maximale (WILKS) : } \sup_{v \in S_n} \eta_n([v])$$

(si $\phi = I_n$ et $\psi = I_p$; ANDERSON [17.])

4.2 Axes principaux de X et composantes principales de tX .

Les axes principaux $[v_k] \subset \{n \times 1\}$ (engendrés par les vecteurs propres de Γ_n) de X et les composantes principales $d_k \in \{n \times 1\}$ de tX engendrent les mêmes sous espaces ; plus précisément

$$\forall k \in]r] : d_k = \lambda_k^{1/2} v_k.$$

$$\text{De même dans } \{p \times 1\} : c_k = \lambda_k^{1/2} w_k.$$

Ces propriétés montrent les rôles symétriques joués par les axes et les composantes principales.

4.3 Recodage des données

Pour déterminer les axes principaux $([v_k])_{1 \leq k \leq r}$, lorsque $\phi \neq I_n$ ou $\psi \neq I_p$, il est toujours possible de se ramener à la recherche d'axes principaux d'un cas $\phi = I_n$ et $\psi = I_p$. Le recodage des données consiste à remplacer X par $\phi_1 X^t \psi_1$ (cf notations) dont les axes sont $(\psi_1 v_k)_{1 \leq k \leq r}$. La k-ème composante principale associée est $\phi_1 X \psi v_k$. Remarquons que si $\phi = I = \phi_1$ cette composante est la même pour tous les choix possibles de ψ_1 .

4.4 Suites complètes de sous-espaces et de matrices d'ajustement

Le sous-espace vectoriel $[V_k]$ de dimension k ($1 \leq k \leq r$) et la matri-

ce $Z_k = \sum_{e=1}^k v_e^t c_e$ de meilleur ajustement de X sont déterminés à partir de la

Pour les différentes valeurs de k , si les valeurs propres non nulles de Γ_p sont simples, on a :

$$(4.1) \quad 1 \leq k < k' \leq r \quad : \quad [v_k] \subset [v_{k'}]$$

et $Z_{k'} = Z_k + Z_{k' - k}$

$$\text{avec } Z_{k' - k} = \sum_{e=k+1}^{k'} v_e t_e.$$

S'il y a des valeurs propres non nulles multiples, il est toujours possible de choisir les sous espaces successifs (ou leurs bases) de sorte que (4.1) soit satisfaite. La suite de sous-espaces vectoriels (resp. de matrices Z_k) est appelée suite complète de sous-espaces (de matrices) d'ajustement : le sous-espace de dimension k de meilleur ajustement est inclus dans le sous-espace de dimension $(k+1)$ de meilleur ajustement.

Pour des applications se reporter à GOOD [7] et DENIAU-OPPENHEIM [5]

BIBLIOGRAPHIE

- [1] ANDERSON (T.W) (1958) *An introduction to multivariate statistical analysis*. Wiley, New-York.
- [2] BENZECRI (J.P) (1973) *L'analyse des correspondances*. Dunod - Paris
- [3] BURT (C.) (1950) The factor Analysis of Qualitative DATA. *Brit. J. Psychol. Stat. Sect., II* pp. 98-121.
- [4] DARROCH (J.N) (1965) An optimal property of principal components *A.M.S.* 36, pp. 1579-1582.
- [5] DENIAU (C.) OPPENHEIM (G.) (1973) A propos de la factorisation des matrices suites complètes conjointes et nuages à pondérations généralisées. (ronéo U.E.R. MLFI Université Paris V). A paraître.
- [6] GABRIEL (K.R) (1971) The biplot graphic display of matrices with application to principal component analysis. *BIOMETRIKA*, 58, 3, pp. 453-466.
- [7] GOOD (I.J) (1969) Some applications of the singular decomposition of a matrix. *Technometrics*, vol 11, n° 44, pp. 823-831.
- [8] KENDALL (H.G) STUART (A.) (1946) *The advanced theory of statistics* (ch. 43) Griffin, London.
- [9] MATHER (K.) (1966) (5ème édition) *Statistical analysis in Biology*. Chapman and Hill.
- [10] OKAMOTO (M.) (1968) *Optimality of principal components* dans KRISHNAIAH P.R., (Ed) *Multivariate Analysis*, Academic Press, 1966.
- [11] RAO (C.R) (1964) The use and interpretation of principal component analysis in applied research, *Sankya, Ser. A*, 26, pp 329-358.
- [12] RAO (C.R) (1965) *Linear statistical inference and its applications*. Wiley, New-York.
- [13] SHETH (J.N) (1969) Using factor analysis to estimate parameters. *J.A.S.A.* 1969-64 pp 808-822

Reçu en Novembre 1974

Université René Descartes
U.E.R. de Mathématiques
Logique formelle et Informatique
12, rue Cujas
75005 PARIS

ETUDE D'UN COEFFICIENT D'INDEPENDANCE
D'UNE VARIABLE ALEATOIRE Y
PAR RAPPORT A UNE VARIABLE ALEATOIRE X.

Jean GUY*

et Jean BASS**

Résumé :

Soit (X, Y) un couple de variables aléatoires. Soit $\omega_{ij} = \Pr(X = x_i \text{ et } Y = y_j)$. On s'intéresse à la manière dont la loi ω_{ij} se rapproche d'une loi d'indépendance. Pour cela, on pose :

$$\omega_{ij} = \mu p_i q_j + (1 - \mu) u_{ij},$$

où $p_i = \sum_j \omega_{ij} = \Pr(X = x_i)$ est la loi marginale relative à X, puis on cherche la plus grande valeur possible μ_0 du coefficient μ ($0 \leq \mu_0 \leq 1$).

$$\text{On montre que } \mu_0 = \sum_j \inf_i \frac{\omega_{ij}}{p_i}$$

Grâce à la définition préalable d'une distance entre matrices stochastiques il apparaît que le rapport $d = \frac{1 - \mu_0}{\mu_0}$ représente la distance de la matrice A

des probabilités conditionnelles :

$$a_{ij} = \frac{\omega_{ij}}{p_i} = \Pr(Y = y_j / X = x_i)$$

à l'ensemble des matrices stochastiques d'indépendance.

On donne enfin quelques exemples d'application à des problèmes concrets:

* U.E.R. Etudes médicales et biologiques de l'Université René Descartes.
45, rue des Saint-Pères. 75006 - Paris.

** U.E.R. 47 de Mathématiques de l'Université Pierre et Marie Curie
Tour 46, 4, place Jussieu, 75280 Paris Cedex 05

44.

I. Enoncé du problème

Soient X et Y deux variables aléatoires prenant respectivement les valeurs x_1, x_2, \dots, x_m et y_1, y_2, \dots, y_n . Appelons :

ω_{ij} la probabilité de $X = x_i, Y = y_j$

p_i la probabilité de $X = x_i$

q_j la probabilité de $Y = y_j$

(les probabilités p_i et q_j étant supposées non nulles).

Si $\omega_{ij} = p_i q_j$, X et Y sont indépendantes.

Pour $m \geq n$, si $\omega_{ij} = p_i \delta_{f(i), j}$ ($f(i)$ étant une application de l'ensemble des indices i sur l'ensemble des indices j), Y devient une fonction de X .

Dans les autres cas, il y a une dépendance stochastique de Y par rapport à X et il convient de chercher un critère simple pour juger la manière dont la dépendance en question se rapproche de l'indépendance.

Le coefficient de corrélation linéaire ρ entre X et Y donne des indications à ce sujet. S'il y a indépendance, il est nul. Mais s'il est nul, on ne peut rien dire. Il peut même arriver que $\rho = 0$ alors que Y est fonction de X , ainsi que cela a lieu si $Y = X^2$, X étant distribuée suivant une loi normale centrée.

Seul le cas où $\rho = \pm 1$ est précis : Y est alors fonction linéaire de X ; réciproquement, $Y = aX + b$ impose $|\rho| = 1$.

Il est donc naturel de chercher d'autres procédés simples pour tester le degré de dépendance de Y par rapport à X . L'objet de ce travail est d'en suggérer un.

Posons :

$$(1) \quad \omega_{ij} = \mu p_i q_j' + (1 - \mu) u_{ij}.$$

p_i est la probabilité pour que $X = x_i$. μ est un nombre compris entre 0 et 1. u_{ij} a, relativement aux indices, les caractères d'une probabilité : $u_{ij} \geq 0$,

$\sum_{ij} u_{ij} = 1$. On suppose enfin que q_j est aussi une probabilité. On vérifie que

toutes ces conditions sont compatibles. On a alors :

$$(2) \quad p_i = \sum_j u_{ij}$$

$$(3) \quad q_j = \mu q'_j + (1 - \mu) \sum_i u_{ij}$$

En général $q'_j \neq q_j$.

Si $\mu = 1$, on a $\omega_{ij} = p_i q'_j = p_i q_j$. Il y a indépendance. Si $\mu = 0$, $\omega_{ij} = u_{ij}$.

Si l'on se donne ω_{ij} , μ n'est pas déterminé. En particulier μ peut prendre la valeur 0. Mais en général μ ne peut pas prendre la valeur 1. Il y a une valeur μ_0 maximale compatible avec la donnée de ω_{ij} . Cette valeur représente jusqu'à un certain point la manière dont la loi ω_{ij} donnée se rapproche de l'indépendance. C'est elle qui va servir à tester le degré de dépendance de Y par rapport à X.

Mais, avant de faire le calcul de μ_0 , nous pouvons interpréter la relation

(1). Introduisons $m + 2$ variables aléatoires indépendantes

Z prenant les valeurs 1 et 0 avec les probabilités μ et $1 - \mu$,

A prenant n valeurs avec les probabilités q'_j .

B_i prenant n valeurs avec les probabilités u_{ij} ($i = 1, 2, \dots, m$).

On choisit la valeur de X. Pour trouver la valeur de Y, une fois X choisie, on commence par tirer au sort la valeur de Z. Si $Z = 1$, on fait une expérience pour déterminer A et l'on prend pour valeur de Y la valeur trouvée pour A, quelle que soit la valeur initiale de X. Si $Z = 0$, on fait une expérience pour déterminer la valeur de la variable B_i dont l'indice correspond à la valeur initiale de X, et l'on prend pour valeur de Y la valeur trouvée pour B_i .

Pratiquement, Z résulte d'une épreuve de Bernoulli (pile ou face généralisée). A est le résultat du tirage dans une urne. Aux B_i correspondent m urnes. On essaie Z. Si $Z = 1$, la valeur de Y résulte du tirage dans la première urne. Si $Z = 0$, on obtient la valeur de Y en tirant une boule de celle des urnes qui correspond à la valeur initialement choisie pour X.

2. Calcul de la valeur maximale μ_0 . Compatibilité

L'équation (1) entraîne :

46.

$$q_j' \leq \frac{\omega_{ij}}{p_i} \quad (\text{quel que soit } i) \text{ et par conséquent :}$$

$$(4) \quad \mu q_j' \leq \inf_i \frac{\omega_{ij}}{p_i}$$

Il en résulte que :

$$(5) \quad \mu \leq \sum_j \left(\inf_i \frac{\omega_{ij}}{p_i} \right).$$

et la valeur maximale de μ est :

$$(6) \quad \mu_0 = \sum_j \left(\inf_i \frac{\omega_{ij}}{p_i} \right)$$

Les valeurs associées des q_j' sont alors données par :

$$(7) \quad \mu_0 q_j' = \inf_i \frac{\omega_{ij}}{p_i}$$

et celles des u_{ij} par (1). Elles sont déterminées, sauf bien entendu si $\mu_0 = 1$.

Montrons que, pour toute valeur de μ telle que $0 < \mu < \mu_0$, le système (1) permet de calculer des q_j' (non complètement déterminés si $\mu < \mu_0$) et des u_{ij} .

On a :

$$(8) \quad \mu q_j' = \frac{\omega_{ij}}{p_i} - (1-\mu) \frac{u_{ij}}{p_i} \leq \frac{\omega_{ij}}{p_i} \leq \inf_i \frac{\omega_{ij}}{p_i}$$

Donnons-nous des $q_j' \geq 0$ satisfaisant à l'inégalité :

$$(9) \quad \mu q_j' \leq \inf_i \frac{\omega_{ij}}{p_i}$$

Nous pouvons en outre demander que $\sum_j q_j' = 1$, puisque :

$$\mu \leq \mu_0 = \sum_j \left[\inf_i \frac{\omega_{ij}}{p_i} \right]$$

Les q_j' étant ainsi choisis, l'égalité (1) donne les u_{ij} . Ils sont bien tels que $\sum_{ij} u_{ij} = 1$. En outre, ils sont bien positifs, car

$$\omega_{ij} - \mu p_i q_j' \geq 0.$$

Le problème a donc en général une infinité de solutions. Il en a une seule si $\mu = \mu_0$. Dans ce cas, on peut préciser le comportement des u_{ij} . Soit i_0 l'indice i pour lequel $\frac{\omega_{ij}}{p_i}$ est minimal. On voit que :

$$(10) \quad u_{i_0 j} = 0.$$

3. Relation entre μ et ρ

On remarque tout d'abord que μ , et en particulier μ_0 , ne dépend que des probabilités des variables X, Y, et non des valeurs prises par ces variables. On ne change donc pas μ , ou μ_0 , si l'on modifie les valeurs X, Y en conservant les probabilités. Au contraire, le coefficient de corrélation linéaire ρ dépend des valeurs prises par X et Y. Mais, pour comparer ρ et μ_0 , on peut choisir ces valeurs comme on le veut. Nous supposons seulement que :

$$EX = EY = 0.$$

A partir de la relation (1), on obtient :

$$\sigma_1^2 = EX^2 = \sum_{ij} \omega_{ij} x_i^2 = \mu \sigma_1^2 + (1 - \mu) \sum_{ij} u_{ij} x_i^2$$

soit encore :

$$(11) \quad \sigma_1^2 = \sum_{ij} u_{ij} x_i^2$$

et l'on a pour σ_2^2 :

48.

$$(12) \quad \sigma_2^2 = \sum_{ij} \omega_{ij} y_j^2 = \mu \sum_{ij} q'_j y_j^2 + (1 - \mu) \sum_{ij} u_{ij} y_j^2$$

La définition même de ρ donne :

$$(13) \quad \rho \sigma_1 \sigma_2 = \sum_{ij} \omega_{ij} x_i y_j = 0 + (1 - \mu) \sum_{ij} u_{ij} x_i y_j.$$

Donc :

$$(14) \quad \rho = \frac{1 - \mu}{\sigma_1 \sigma_2} \sum_{ij} u_{ij} x_i y_j$$

Posons

$$r = \frac{\sum_{ij} u_{ij} x_i y_j}{\sqrt{\sum_{ij} u_{ij} x_i^2} \cdot \sqrt{\sum_{ij} u_{ij} y_j^2}}$$

r n'est pas en général un coefficient de corrélation car, pour la loi u_{ij} ,

Y n'est pas de valeur moyenne nulle. Mais on a bien entendu :

$$|r| \leq 1.$$

On obtient :

$$(15) \quad \rho = \frac{(1 - \mu) r}{\sigma_2} \sqrt{\sum_{ij} u_{ij} y_j^2}$$

et d'après (12)

$$(16) \quad \sqrt{1 - \mu} \sqrt{\sum_{ij} u_{ij} y_j^2} \leq \sigma_2.$$

Donc :

$$(17) \quad \rho \leq \sqrt{1 - \mu}$$

Cette dernière inégalité, valable pour tout μ , l'est en particulier pour μ_0 .

4. Exemple du couple de variables aléatoires à deux valeurs

Supposons $m = n = 2$. On va a priori choisir les deux valeurs possibles de X et de Y pour que $EX = 0$, $EY = 0$. Alors elles sont définies à un facteur près, qui n'a pas d'influence sur les calculs et les résultats.

Nous désignerons par p , $1 - p$ les probabilités des valeurs x_1 et x_2 de X, par q , $1 - q$ celles des valeurs y_1 et y_2 de Y, par λ la probabilité du couple (x_1, y_1) . A un facteur près, on a nécessairement :

$$x_1 = 1 - p \qquad x_2 = -p$$

$$y_1 = 1 - q \qquad y_2 = -q$$

ce qui permet de construire le tableau usuel suivant des probabilités des couples de valeurs et des probabilités marginales :

X Y	1 - p	- p	
1 - q	λ	$q - \lambda$	q
- q	$p - \lambda$	$1 + \lambda - p - q$	$1 - q$
	p	$1 - p$	1

On a :

$$(18) \quad p + q - 1 \leq \lambda \leq \inf(p, q).$$

Il y a indépendance si $\lambda = pq$.

Il n'est pas en général possible de choisir λ , une fois p et q donnés, pour que Y soit fonction de X. Il faut pour cela que $p = q$. Alors, si $\lambda = p = q$, le tableau ci-dessus devient :

$$\begin{pmatrix} p & 0 \\ 0 & 1 - p \end{pmatrix}$$

50.

On trouve facilement :

$$\begin{aligned} EX^2 &= p(1-p) & EY^2 &= q(1-q) \\ EXY &= \lambda - pq. \end{aligned}$$

Le coefficient de corrélation entre X et Y est :

$$(19) \quad \rho = \frac{\lambda - pq}{\sqrt{p(1-p)} \cdot \sqrt{q(1-q)}}$$

Pour calculer μ , on écrit :

	$\lambda = \mu pr + (1 - \mu) u_{11}$	
	$p - \lambda = \mu p(1 - r) + (1 - \mu) u_{12}$	
	$q - \lambda = \mu(1 - p)r + (1 - \mu) u_{21}$	
	$1 + \lambda - p - q = \mu(1 - p)(1 - r) + (1 - \mu) u_{22}$	
On a :		

$$\mu r = \inf\left(\frac{\lambda}{p}, \frac{q - \lambda}{1 - p}\right), \quad \mu(1 - r) = \inf\left(\frac{p - \lambda}{p}, \frac{1 + \lambda - p - q}{1 - p}\right)$$

La valeur de r dépend du signe de $\lambda - pq$. La valeur maximale μ_0 de μ est dans tous les cas égale à $\mu_0 = 1 - \frac{|\lambda - pq|}{p(1-p)}$, ce qui s'écrit aussi :

$$(20) \quad 1 - \mu_0 = \frac{|\lambda - pq|}{p(1-p)}$$

Il s'agit du μ_0 correspondant aux probabilités conditionnelles de Y une fois X choisi. Si l'on permute X et Y, on trouve un μ'_0 tel que :

$$(21) \quad 1 - \mu'_0 = \frac{|\lambda - pq|}{q(1-q)}$$

On constate que :

$$(22) \quad (1 - \mu_0)(1 - \mu'_0) = \rho^2$$

On vérifie que μ_0 et μ'_0 sont compris entre 0 et 1 et que :

$$(23) \quad \rho^2 \leq \inf(1 - \mu_0, 1 - \mu'_0)$$

5. **Interprétation du coefficient μ_0 à l'aide de la distance de deux matrices stochastiques.**

Nous poserons $a_{ij} = \frac{\omega_{ij}}{p_i}$. a_{ij} est la probabilité conditionnelle de $Y = y_j$, lorsque $X = x_i$. La matrice $A = (a_{ij})$ est une matrice stochastique :

$$a_{ij} \geq 0, \quad \sum_j a_{ij} = 1.$$

Appelons E_{mn} l'ensemble des matrices stochastiques A à m lignes et n colonnes.

Nous allons maintenant définir une distance $d(A, B)$ entre deux matrices $A, B \in E_{mn}$ par :

$$(24) \quad d(A, B) = \frac{1}{m} \sum_{ij} \left| \frac{a_{ij} - b_{ij}}{\mu_{A0} \mu_{B0}} \right|,$$

μ_{A0} et μ_{B0} étant les deux constantes entièrement définies à partir des matrices A et B, suivant :

$$(25) \quad \mu_{A0} = \sum_j (\inf_i a_{ij}), \quad \mu_{B0} = \sum_j (\inf_i b_{ij}).$$

Vérifions tout d'abord que $d(A, B)$ est bien une distance.

a) La relation de définition (24) demeure inchangée par permutation de A et de B, d'où $d(A, B) = d(B, A)$. On vérifie de plus immédiatement que $d(A, A) = 0$.

b) Une valeur nulle de $d(A, B)$ entraîne l'identité des matrices A et B. Remarquons tout d'abord que $d(A, B) = 0 \Rightarrow \mu_{A0} \neq 0$ et $\mu_{B0} \neq 0$ (la distance $d(A, B)$ devient infiniment grande si une et une seulement des deux constantes μ_{A0} et μ_{B0} est nulle ; elle n'est plus définie si $\mu_{A0} = \mu_{B0} = 0$). Par suite $d(A, B) = 0$ entraîne :

52.

$$(26) \quad \frac{a_{ij}}{\mu_{A_0}} = \frac{b_{ij}}{\mu_{B_0}} \quad (\forall i, j) \quad \text{c'est-à-dire} \quad a_{ij} = \frac{\mu_{A_0}}{\mu_{B_0}} b_{ij}.$$

Par sommation sur j de la dernière égalité, il vient :

$$(27) \quad 1 = \frac{\mu_{A_0}}{\mu_{B_0}}, \text{ d'où il résulte que } a_{ij} = b_{ij} \quad (\forall i, j).$$

c) L'inégalité triangulaire

$$(28) \quad d(A, C) \leq d(A, B) + d(B, C)$$

est également valable. Ecrivons :

$$d(A, B) = d'(A', B'), \quad d(A, C) = d'(A', C'), \quad d(B, C) = d'(B', C'),$$

où A', B' et C' représentent les matrices (non stochastiques cette fois) définies par :

$$A' = \frac{1}{\mu_{A_0}} A, \quad B' = \frac{1}{\mu_{B_0}} B, \quad C' = \frac{1}{\mu_{C_0}} C,$$

On a :

$$(29) \quad d'(A', B') = \frac{1}{m} \sum_{ij} |a'_{ij} - b'_{ij}|$$

et des formules analogues. Or (29) définit l'une des distances usuelles entre deux matrices de dimensions (m, n), distance pour laquelle on a bien :

$$d'(A', C') \leq d'(A', B') + d'(B', C').$$

Lorsque la matrice B devient une matrice stochastique d'indépendance (les m lignes sont alors identiques), on peut poser $b_{ij} = b_j (\forall i)$. Donc :

$$(30) \quad \mu_{B_0} = \sum_j \inf_i (b_j) = \sum_j b_j = 1.$$

Prenons en particulier pour matrice B la matrice d'indépendance Q' dont les éléments q'_j découlent de ceux de la matrice A par la résolution des équations

$$(31) \quad a_{ij} = \mu_{A_0} q'_j + (1 - \mu_{A_0}) v_{ij}$$

(v_{ij} étant automatiquement les éléments d'une nouvelle matrice stochastique V).
On obtient :

$$\begin{aligned}
 (32) \quad d(A, Q') &= \frac{1}{m} \sum_{ij} \left| \frac{a_{ij}}{\mu_{Ao}} - q'_j \right| \\
 &= \frac{1}{m} \sum_{ij} \left| q'_j + \frac{1 - \mu_{Ao}}{\mu_{Ao}} v_{ij} - q'_j \right| \\
 &= \frac{1}{m} \frac{1 - \mu_{Ao}}{\mu_{Ao}} \sum_{ij} v_{ij} = \frac{1}{m} \frac{1 - \mu_{Ao}}{\mu_{Ao}} \sum_i 1 \\
 &= \frac{1 - \mu_{Ao}}{\mu_{Ao}}
 \end{aligned}$$

$\frac{1 - \mu_{Ao}}{\mu_{Ao}}$ est donc la distance de la matrice A à la matrice d'indépendance qui

lui est associée par la formule (31).

Nous voulons montrer que, si b_j représente les éléments d'une matrice stochastique d'indépendance quelconque, on a :

$$(33) \quad \frac{1}{m} \sum_{ij} \left| \frac{a_{ij}}{\mu_{Ao}} - b_j \right| \geq \frac{1 - \mu_{Ao}}{\mu_{Ao}}$$

Or, par l'hypothèse,

$$a_{ij} = \mu_{Ao} q'_j + (1 - \mu_{Ao}) v_{ij}$$

Nous poserons $b_j = q'_j + r_j$.

Le premier membre de (33) s'écrit donc :

$$(34) \quad \frac{1}{m} \sum_{ij} \left| \frac{1 - \mu_{Ao}}{\mu_{Ao}} v_{ij} - r_j \right|$$

54.

$1 - \mu_{Ao}$ est différent de 0. Nous savons que certains des v_{ij} sont nuls. Mais ils ne le sont pas tous. Nous allons d'abord supposer que la matrice $B = (b_j)$ (à n lignes identiques) est suffisamment voisine de la matrice $Q' = q_j'$ pour que, lorsque $v_{ij} \neq 0$, on ait :

$$(35) \quad \frac{1 - \mu_{Ao}}{\mu_{Ao}} v_{ij} - r_j > 0.$$

Dans (34), il y a donc des termes qui se réduisent à $|r_j|$ et d'autres qui sont représentés par (35), au facteur $\frac{1}{m}$ près. i étant choisi, faisons dans (34) la sommation par rapport à j . Il y a deux sortes de termes :

$$(36) \quad \sum_{j'} \left(\frac{1 - \mu_{Ao}}{\mu_{Ao}} v_{ij'} - r_{j'} \right) \quad (j' \text{ tel que } v_{ij'} \neq 0)$$

$$(37) \quad \sum_{j''} |r_{j''}| \quad (j'' \text{ tel que } v_{ij''} = 0)$$

Or (36) est la somme de :

$$(38) \quad \sum_j \frac{1 - \mu_{Ao}}{\mu_{Ao}} \quad (\text{pour toutes les valeurs de } j)$$

et de :

$$(39) \quad - \sum_{j'} r_{j'}$$

Au total, en ajoutant (36) et (37), on obtient :

$$(40) \quad \sum_j \frac{1 - \mu_{Ao}}{\mu_{Ao}} v_{ij} + \sum_{j''} |r_{j''}| - \sum_{j'} r_{j'}$$

Mais $\sum_j r_j = 0$. Donc $\sum_{j'} r_{j'} = - \sum_{j''} r_{j''}$ et l'on obtient :

$$(41) \quad \sum_j \frac{1 - \mu_{Ao}}{\mu_{Ao}} v_{ij} + \sum_{j''} |r_{j''}| + \sum_{j''} r_{j''}$$

soit :

$$(42) \quad \frac{1 - \mu_{Ao}}{\mu_{Ao}} + H,$$

où H est une quantité ≥ 0 . On somme enfin par rapport à i, on divise par m, et l'on voit que (34) est la somme de $\frac{1 - \mu_{Ao}}{\mu_{Ao}}$ et d'une quantité ≥ 0 . Cela prouve l'inégalité (33).

La démonstration
prouve que $\frac{1 - \mu_{Ao}}{\mu_{Ao}}$ est

un minimum relatif pour la distance entre la matrice stochastique A et l'ensemble E des matrices d'indépendance. Mais E est un ensemble convexe. Si donc la distance entre le point A et l'ensemble E admet un minimum relatif au point Q', c'est aussi le minimum absolu. En effet, s'il existait un point B de E tel que :

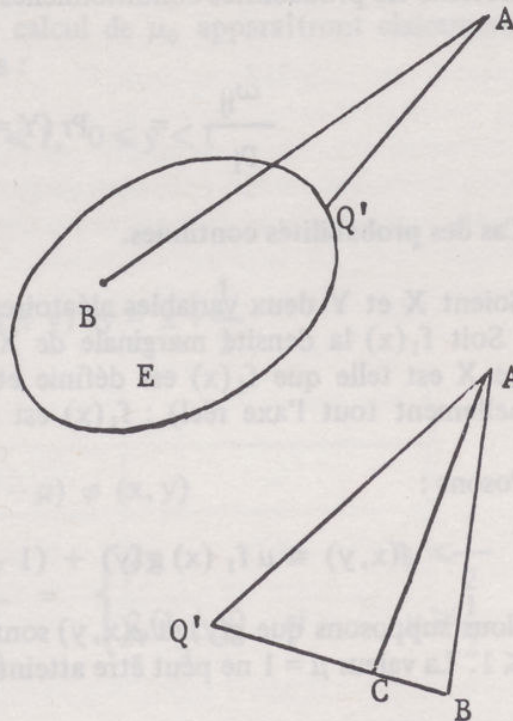


Fig. 1

$$d(A, B) < d(A, Q'),$$

on aurait :

$$d(A, C) < d(A, Q')$$

pour tout C du segment BQ'. Mais $C \in E$ (cf. fig. 1). Il y aurait donc, aussi près qu'on le veut de Q', un point C tel que $d(A, C) < d(A, Q')$ et d(A, Q') ne serait pas un minimum relatif.

56.

On peut conclure de (33) que :

$$d(A, Q') = \frac{1 - \mu_{Ao}}{\mu_{Ao}}$$

est la distance de la matrice stochastique A à l'ensemble des matrices stochastiques d'indépendance. La matrice Q', complètement définie par les équations (31), est la matrice d'indépendance la plus proche de la matrice stochastique A qui représente les probabilités conditionnelles :

$$\frac{\omega_{ij}}{p_i} = \Pr(Y = y_j / X = x_i).$$

6. Cas des probabilités continues.

Soient X et Y deux variables aléatoires ayant une densité de probabilité $f(x, y)$. Soit $f_1(x)$ la densité marginale de X. Nous supposons que la variable aléatoire X est telle que $f_1(x)$ est définie et ne s'annule pas sur un intervalle (éventuellement tout l'axe réel) ; $f_1(x)$ est nulle en dehors de cet intervalle.

Posons :

$$(43) \quad f(x, y) = \mu f_1(x) g(y) + (1 - \mu) \varphi(x, y)$$

Nous supposons que $g(y)$ et $\varphi(x, y)$ sont des densités de probabilité et que $0 \leq \mu \leq 1$. La valeur $\mu = 1$ ne peut être atteinte que si :

$$(44) \quad f(x, y) = f_1(x) g(y)$$

Il y a alors indépendance et $g(y)$ coïncide avec la densité de probabilité marginale $f_2(y)$ de Y.

Dans le cas général, on a :

$$(45) \quad \mu f_1(x) g(y) \leq f(x, y)$$

On en déduit :

$$(46) \quad \mu g(y) \leq \frac{f(x, y)}{f_1(x)}$$

(densité de probabilité conditionnelle de Y pour $X = x$). Par suite :

$$(47) \quad \mu g(y) \leq \inf_x \frac{f(x, y)}{f_1(x)}$$

La valeur maximale μ_0 de μ est telle que :

$$(48) \quad \mu_0 g(y) = \inf_x \frac{f(x, y)}{f_1(x)}$$

On voit que :

$$(49) \quad \mu_0 = \int_{-\infty}^{+\infty} \inf_x \frac{f(x, y)}{f_1(x)} dy.$$

Exemple : Les modalités de calcul de μ_0 apparaîtront clairement sur le très simple exemple suivant : Posons :

$$f(x, y) = x + y \text{ pour } 0 < x \leq 1, 0 \leq y < 1$$

$$f(x, y) = 0 \text{ ailleurs.}$$

$$\text{On a } f_1(x) = \int_0^1 (x + y) dy = x + \frac{1}{2}$$

On pose :

$$x + y = \mu(x + \frac{1}{2}) g(y) + (1 - \mu) \varphi(x, y)$$

$$g(y) \mu_0 = \inf_{0 \leq x \leq 1} \frac{x + y}{x + \frac{1}{2}} = \begin{cases} 2y & \text{si } y \leq \frac{1}{2} \\ \frac{2(1+y)}{3} & \text{si } y \geq \frac{1}{2} \end{cases}$$

Donc :

$$\mu_0 = \int_0^{\frac{1}{2}} 2y dy + \int_{\frac{1}{2}}^1 \frac{2(1+y)}{3} dy = \frac{5}{6}$$

7. Exemples d'applications.

Dès qu'une variable aléatoire Y est stochastiquement dépendante d'une autre variable aléatoire X, le coefficient d'indépendance μ_0 défini par (6) et la loi d'indépendance q_j^i la plus proche de la loi conditionnelle $\frac{\omega_{ij}}{P_i}$

des caractéristiques intéressantes pour ce couple de variables. Les quelques exemples simples traités ci-après sont destinés à mieux préciser les informations

58.

liées à la connaissance de μ_0 et de la loi q_j .

a) Etude de la corrélation existant entre les sexes des jumeaux.

A partir des données statistiques concernant les grossesses gémellaires, il est possible d'estimer les diverses probabilités pour que les deux jumeaux soient, dans l'ordre de leur naissance, garçon-garçon (ω_{GG}), garçon-fille (ω_{GF}), fille-garçon (ω_{FG}) et fille-fille (ω_{FF}). On peut considérer comme valables les valeurs numériques :

$$\begin{aligned} \omega_{GG} &= 0,329, & \omega_{GF} &= \omega_{FG} = 0,179. \\ \omega_{FF} &= 0,313. \end{aligned}$$

qui conduisent au tableau suivant représentant la matrice stochastique A (probabilités conditionnelle pour que le second jumeau soit un garçon ou une fille lorsque le sexe du premier jumeau est connu).

	2e jumeau	G	F
1er jumeau			
G		0,648	0,352
F		0,364	0,636

Par suite $\mu_0 = 0,364 + 0,352 = 0,716$ et la distance à la loi d'indépendance la plus proche est

$$d = \frac{1}{\mu_0} - 1 = 0,397.$$

Remarquons ici que l'égalité $\omega_{GF} = \omega_{FG}$ entraîne $\rho = 1 - \mu_0 = 0,284$

La loi q_j apparaît ensuite nettement si nous décomposons la matrice A conformément à (31), soit :

$$\begin{aligned} (50) \quad A &= \begin{pmatrix} 0,648 & 0,352 \\ 0,364 & 0,636 \end{pmatrix} \\ &= \mu_0 \begin{pmatrix} 0,508 & 0,492 \\ 0,508 & 0,492 \end{pmatrix} + (1 - \mu_0) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Le complément de μ_0 à l'unité ($1 - \mu_0 = 0,284$) avait déjà été interprété par E. Borel [1]. Ce complément précise le nombre de fois où le sexe du second enfant est complètement déterminé par le sexe du premier (jumeaux «vrais»): l'identité des sexes, due alors à l'identité des génotypes, découle analytiquement

de la présence de la matrice unité dans le terme $(1 - \mu_0)$ $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

de la décomposition (50).

b) Valeurs de μ_0 dans les phénomènes d'hérédité liée au sexe.

Dans le cas des espèces animales analogues à l'espèce humaine du point de vue du déterminisme sexuel (chromosomes sexuels XX chez la femelle et XY chez le mâle), considérons un gène porté par le chromosome X et susceptible d'exister sous deux états allèles G et g : G (dominant chez les femelles) conduit à un phénotype normal cependant que le génotype g pour le mâle ou gg pour la femelle conduit à un phénotype taré. Si η est la proportion de chromosomes X porteurs de g pour l'ensemble de la population (les croisements étant effectués au hasard) et si α et β sont les proportions de mâles et de femelles, on obtient sans difficultés le tableau suivant (matrice stochastique A) pour les probabilités conditionnelles qu'un sujet soit mâle ou femelle lorsque nous savons que son phénotype est normal (N) ou taré (T).

Phénotype \ Sexe	δ	♀
	N	$\frac{\alpha(1-\eta)}{1-\alpha\eta-\beta\eta^2}$
T	$\frac{\alpha}{\alpha+\beta\eta}$	$\frac{\beta\eta}{\alpha+\beta\eta}$

Pour $\alpha = \beta = \frac{1}{2}$, nous trouvons :

$$(51) \quad \mu_0 = \frac{1}{2+\eta} + \frac{\eta}{1+\eta} = \frac{1+3\eta+\eta^2}{(2+\eta)(1+\eta)}$$

Les courbes représentant μ_0 et $d = \frac{1}{\mu_0} - 1$ en fonction de η sont portées sur la fig. 2. On constate notamment que $0,5 < \mu_0 < 0,833$, les limites indiquées étant respectivement atteintes de manière asymptotique pour η tendant vers zéro ou vers l'unité. Deux statistiques effectuées séparément sur des sujets

normaux et sur des sujets tarés doivent donc permettre d'atteindre μ_0 : si la valeur trouvée pour μ_0 diffère significativement de l'unité, on pourra conclure à l'existence d'une hérédité liée au sexe et un tel phénomène sera d'autant plus facile à mettre en évidence par l'intermédiaire de μ_0 que la tare sera moins répandue, puisque η très faible conduit à la valeur minimale de μ_0 .

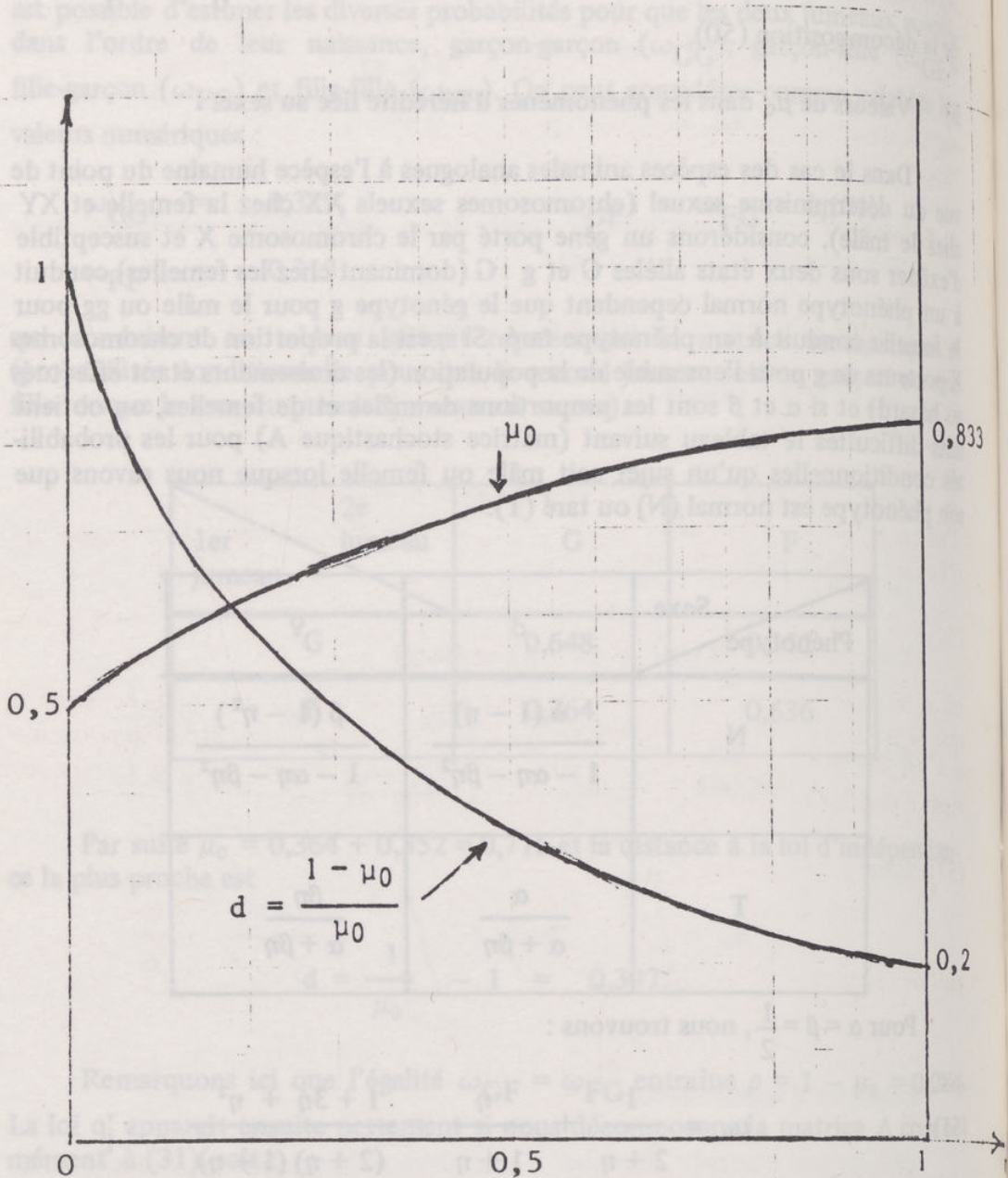


Fig. 2

c) Possibilité de tester les phénomènes de pléiotropie par un coefficient d'indépendance.

Toujours en biologie, il y a pléiotropie lorsqu'un gène comportant plusieurs états allèles participe simultanément à la réalisation de plusieurs caractères phénotypiques. Il conviendra de former ici la matrice stochastique A décrivant les probabilités conditionnelles des phénotypes possibles pour le deuxième caractère lorsque les phénotypes relatifs au premier caractère sont connus. Comme il ne peut y avoir indépendance totale entre les deux séries de phénotypes lorsqu'une liaison existe par suite de la pléiotropie, une différence significative de μ_0 (coefficient d'indépendance du deuxième caractère par rapport au premier) vis-à-vis de l'unité signalera un tel phénomène.

d) Emploi de μ_0 dans certains problèmes de mécanique quantique.

Notre dernier exemple concernera les distributions électroniques dans les atomes et les molécules. Pour l'état électronique fondamental de l'atome d'hélium, une approximation usuelle est celle du produit simple de deux orbitales atomiques identiques afin de représenter les deux électrons 1s. C'est là un modèle d'indépendance totale conduisant à $\mu_0 = 1$ lorsque nous étudions les densités de probabilité de présence dans tout l'espace du 2^e électron en fonction de la position du premier électron, supposée préalablement fixée. Toutefois, d'autres types de fonctions propres approchées sont possibles, comme celle d'Eckart et Hylleraas [2], soit :

$$(52) \quad \psi(1, 2) = N [\exp(-\alpha r_1 - \beta r_2) + \exp(-\beta r_1 - \alpha r_2)],$$

où r_1 et r_2 représentent les distances des électrons 1 et 2 au noyau cependant que N est le facteur usuel de normalisation. L'ajustement des paramètres donne $\alpha = 2,15$ et $\beta = 1,19$ en unités atomiques de Hartree. Pour le calcul de μ_0 , il convient d'évaluer les densités de probabilités conditionnelles :

$$(53) \quad f(2/1) = \frac{\psi^2(1, 2)}{\int \psi^2(1, 2) d\tau_2}$$

puis de calculer :

$$(54) \quad \mu_0 = \int \inf_{(1)} [f(2/1)] d\tau_2$$

On trouve de cette manière $\mu_0 = 0,737$ pour la fonction propre approchée (41).

Dans cette application à des variables aléatoires de nature continue (X et Y caractérisent les positions des électrons numérotés 1 et 2), nous vérifions

que les électrons du modèle proposé par Eckart et Hylleraas ne sont plus totalement indépendants bien que μ_0 reste relativement élevé. Notons d'ailleurs que la fonction propre rigoureuse devrait conduire à $\mu_0 = 0$ car $\inf [f(2/1)]$ (1)

reste constamment nul pour une bonne fonction propre, la répulsion coulombienne interdisant à deux électrons d'être simultanément présents dans un même élément de volume.

REFERENCES

- [1] BOREL (E.) *Eléments de la théorie des probabilités*, p. 155 Hermann édit. (1924)
- [2] cf. PAULING (L.) *Introduction to quantum mechanics* p. 224. Mc Graw Hill édit. (1935).

Reçu en Juin 1973.

Etude d'un coefficient d'indépendance d'une variable stochastique Y par rapport
à une variable stochastique X

1973

REFERENCES

1. H. DEBIEVE, *Revue de Statistique*, 1972, 29, 1, 1-10.
2. H. DEBIEVE, *Revue de Statistique*, 1973, 30, 1, 1-10.
3. H. DEBIEVE, *Revue de Statistique*, 1973, 30, 2, 1-10.
4. H. DEBIEVE, *Revue de Statistique*, 1973, 30, 3, 1-10.
5. H. DEBIEVE, *Revue de Statistique*, 1973, 30, 4, 1-10.

Reçu en Juin 1973.