P. CAZES

Problème sur l'analyse discriminante

Les cahiers de l'analyse des données, tome 3, nº 1 (1978),
p. 103-108

<http://www.numdam.org/item?id=CAD_1978__3_1_103_0>

Les Cahiers de l'Analyse des Données
Vol. III 1978 - n° 1 p. 103 108

PROBLÈME SUR L'ANALYSE DISCRIMINANTE [CLOISON OPTIMALE]

Solution par P. Cazes [1]

Le but du présent problème (*) est de montrer que les méthodes usuelles d'analyse discriminante, fondées sur la métrique d'inertie, peuvent ne pas fournir la cloison optimale pour séparer deux distributions de probabilité (cf § 2.5 à 2.7).

1. Enoncé du problème:

Dans le plan rapporté à deux axes de coordonnées ox^1, ox^2, on considère deux carrés A et B dont les côtés sont parallèles aux axes:

carré A : centre $\alpha = (\alpha^1, \alpha^2)$; côté 2a ; masse p ;

carré B : centre $\beta = (\beta^1, \beta^2)$; côté 2b ; masse $q = 1 - p$;

qui ont tous deux une densité uniforme, ρ pour le carré A et η pour le carré B.

1.1 A quelles conditions doivent satisfaire les données (α^1, α^2), a, p (β^1, β^2), b pour que le centre de gravité du système formé des deux carrés A et B soit à l'origine ?

1.2 Sous cette hypothèse écrire la forme quadratique d'inertie σ du système A + B relativement à l'origine (on calculera les moments d'inertie σ^{11}, σ^{22} et le produit d'inertie σ^{12}).

1.3 A quelle condition deux formes linéaires u et v

$$u(x) = u_1\, x^1 + u_2\, x^2$$

$$v(x) = v_1\, x^1 + v_2\, x^2$$

sont-elles conjuguées, (i.e. perpendiculaires) par rapport à la métrique d'inertie du nuage A + B .

1.4 Quelle est la droite Δ passant par l'origine, perpendiculaire (pour la métrique d'inertie de A + B) à la droite $\alpha\beta$ joignant les centres des deux carrés A et B.

1.5 On suppose que la droite Δ est utilisée comme cloison discriminante entre A et B. Quelle est alors la masse de la partie du carré A située par rapport à Δ de l'autre côté que le centre α ? De même, quelle est la masse de la partie du carré B située par rapport à Δ de l'autre côté que le centre β ? Sans essayer de trouver une formule générale, on pourra se borner au cas particulier suivant qui ne dépend que du paramètre (h/a).

(*) *Ce problème a été proposé aux étudiants du D.E.A. de statistique de l'Université Pierre et Marie Curie, en Mars 1974.*

(1) *Maître-Assistant I.S.U.P. - Laboratoire de Statistique - Université Pierre et Marie Curie, Paris.*

$$\alpha^1 = a = b = -\beta^1$$

$$\alpha^2 = -\beta^2 = h$$

$$p = q = 1/2$$

1.6 Existe-t-il dans le cas particulier ci-dessus, pour certaines valeurs de h/a une cloison Δ' parallèle à Δ et réalisant une meilleure discrimination que Δ ?

1.7 Est-il possible que la meilleure discrimination entre A et B soit fournie par une droite non-parallèle à Δ ?

1.8 Reprendre la question 1.5 pour établir une formule générale.

2 *Solution du problème*

2.1 Le centre de gravité des deux carrés est à l'origine si :

$$p\,\alpha^i + q\,\beta^i = 0 \qquad (i = 1 \text{ et } 2) \tag{1}$$

2.2 Soit μ la mesure associée à l'ensemble des deux carrés ; on a puisque le centre de gravité est à l'origine :

$$\sigma^{ij} = \int_{A+B} x^i\, x^j\, d\mu(x^1,x^2)$$

soit en appliquant deux fois le théorème de Huyghens, et en se rappelant que pour un carré de côté $2l$, de densité uniforme, et centré en O on a : $(1/4\,l^2)\int_{-l}^{l}\int_{-l}^{l}(x^i)(x^j)dx^i\,dx^j = l^2\delta_i^{\,j}/3$ (i.e. moment d'inertie du carré par rapport à un axe passant par son centre égal à $l^2/3$, et produits d'inertie nuls) :

$$\sigma^{ij} = (p/4a^2)\int_A (x^i-\alpha^i)(x^j-\alpha^j)dx^i\,dx^j + p\,\alpha^i\,\alpha^j +$$

$$(q/4b^2)\int_B (x^i-\beta^i)(x^j-\beta^j)dx^i\,dx^j + q\,\beta^i\,\beta^j$$

$$= (p\,a^2 + q\,b^2)\delta_i^{\,j}/3 + p\,\alpha^i\,\alpha^j + q\,\beta^i\,\beta^j$$

soit encore en tenant compte de (1) :

$$\sigma^{ij} = (p\,a^2 + q\,b^2)\;\delta_i^{\,j}/3 \;+\; p\,\alpha^i\,\alpha^j/q = (\sigma_0)^{ij} + (\sigma_1)^{ij} \tag{2}$$

avec : $(\sigma_0)^{ij} = (p\,a^2 + q\,b^2)\;\delta_i^{\,j}/3$: terme général de la matrice variance intraclasse σ_0.

$(\sigma_1)^{ij} = p\,\alpha^i\,\alpha^j + q\,\beta^i\,\beta^j = p\,\alpha^i\,\alpha^j/q$: terme général de la matrice variance intraclasse σ_1.

On aurait bien sûr pu calculer σ en appliquant la formule $\sigma = \sigma_0 + \sigma_1$.

Remarquons que σ_0 est à un coefficient près la matrice unité, ceci étant dû à la symétrie sphérique des carrés A et B ; par ailleurs σ_1 est de rang 1, puisque c'est la matrice d'inertie associée aux points α et β respectivement affectés des masses p et q.

2.3 u et v sont conjugués si

$$\sigma(u,v) = \sigma^{11} u_1 v_1 + \sigma^{12}(u_1 v_2 + u_2 v_1) + \sigma^{22} u_2 v_2 = 0$$

2.4 L'équation de la droite Δ perpendiculaire pour la métrique d'inertie σ^{-1} à $\alpha\beta$ est :

$$\sigma^{-1}(\alpha\beta\ ,\ x) = 0$$

or $$\sigma^{-1} = \frac{1}{\sigma^{11}\sigma^{22} - (\sigma^{12})^2} \begin{pmatrix} \sigma^{22} & -\sigma^{12} \\ -\sigma^{12} & \sigma^{11} \end{pmatrix}$$

L'équation de la droite Δ s'écrit donc

$$\sigma^{22}(\beta^1 - \alpha^1)\, x^1 + \sigma^{11}(\beta^2 - \alpha^2)\, x^2 - \sigma^{12}((\beta_1{}^1 - \alpha^1)\, x^2 + (\beta^2 - \alpha^2)\, x^1) = 0$$

ou encore puisque d'après (1), $\beta^i - \alpha^i = -\alpha^i/q$:

$$\sigma^{22}\, \alpha^1\, x^1 + \sigma^{11}\, \alpha^2\, x^2 - \sigma^{12}(\alpha^1\, x^2 + \alpha^2\, x^1) = 0$$

ce qui s'écrit d'après (2), et en posant $c^2 = (p\,a^2 + q\,b^2)/3$

$$(c^2 + p(\alpha^2)^2/q)\,\alpha^1 x^1 + (c^2 + p(\alpha^1)^2/q)\,\alpha^2 x^2 - (p/q)\alpha^1\alpha^2(\alpha^1 x^2 + \alpha^2 x^1) = 0$$

soit $$\alpha^1\, x^1 + \alpha^2\, x^2 = 0 \qquad (3)$$

La droite Δ est donc orthogonale au sens usuel à $O\alpha$, i.e. à la droite $\alpha\,\beta$ joignant les deux centres de gravité.

On aurait pu établir ce résultat, directement, sans aucun calcul ; en effet, Δ a pour équation :

$$\Delta = \{x \in E \mid \sigma^{-1}(x, \alpha\beta) = 0\}$$

où l'on a désigné par E le plan des 2 carrés.

Du fait de l'isomorphisme induit par σ^{-1} entre E et E^*, on a en posant $u = \sigma^{-1}(x)$, ou $x = \sigma(u)$, et en désignant par Δ^* l'isomorphe de Δ dans E^* :

$$\Delta^* = \sigma^{-1}\ (\Delta) = \{\,u \in E^* \mid u(\alpha\beta\,\} = 0$$

$$\Delta = \sigma(\Delta^*)$$

Or sur Δ^*, σ et σ_0 coïncident puisque la restriction de σ_1 à Δ^* est nulle ; on a donc :

$$\Delta = \sigma(\Delta^*) = \sigma_0\,(\Delta^*) = \{\sigma_0(u) \mid u(\alpha\beta) = 0\} = \{x \mid \sigma_0^{-1}(x, \alpha\beta) = 0\}$$

Δ est donc orthogonal à $\alpha\beta$ pour la métrique σ_0^{-1}.

Comme σ_0 est à un coefficient près la métrique unité, Δ est orthogonal à $\alpha\,\beta$ au sens de la métrique usuelle.

2.5 Nous supposerons à partir de ce §2.5 que les deux carrés A et B ont leurs côtés parallèles aux axes, les résultats trouvés aux quatres premières questions étant valables sans cette condition.

Ayant : $\alpha^1 = a = b = -\beta^1$; $\alpha^2 = -\beta^2 = h$; $p = q = 1/2$, on vérifie bien que le centre de gravité est à l'origine.

D'après (3), Δ a pour équation : $a x^1 + h x^2 = 0$

Dans le cas particulier envisagé, l'axe Ox^2 est le support d'un côté du carré A et d'un côté du carré B. Deux cas sont à envisager suivant que h est plus grand ou plus petit que a (cf figure 1).

(1) : $h > a$: dans ce cas là la droite Δ ne coupe pas les carrés A et B qu'elle sépare parfaitement.

(2) : $h \leq a$: dans ce cas, Δ coupe A au point M_0 de coordonnées $x_0^1 = (h/a)(a-h)$, $x_0^2 = -(a-h)$; si I désigne le point de coordonnées $(0, -(a-h))$ la portion de A opposée à α par rapport à Δ est le triangle rectangle $O I M_0$ de surface $x_0^1\ |x_0^2|/2 = h(a-h)^2/2a$ et la masse associée vaut donc $(p/4a^2)\, h(a-h)^2/2a = h(a-h)^2/8a^3$. La figure étant symétrique par rapport à l'origine, c'est également la masse de B située de l'autre côté de β par rapport à Δ.

2.6 Dans le cas (1), toute droite Δ' parallèle à Δ, et située entre les points $I(0, h-a)$ et $J(0,-(h-a))$ assure comme Δ une parfaite discrimination.

Dans le cas (2), il est facile de voir que pour toute droite Δ' parallèle à Δ , la discrimination est moins bonne, la masse, (ou la surface) des points mal classés augmentant.

Dans le cas de la figure 2 ci-dessous, la surface des mal classés augmente de la différence : aire du trapèze $O M_0 L K$ - aire du trapèze $O N_0 J' K$, différence qui est positive, ces deux trapèzes ayant des hauteurs égales, même longueur pour le premier côté de base, le second côté de base du premier LK étant plus grand que le second côté de base du second KJ'.

2.7 Dans le cas (1), toute droite de pente négative, coupant Ox^2 entre I et J assure une parfaite discrimination comme Δ.

Dans le cas (2), la droite Ox^2 assure une parfaite discrimination entre A et B contrairement à Δ. Ceci est dû au fait qu'en analyse factorielle discriminante, on fait la discrimination en prenant comme critère, le critère de la combinaison linéaire de variance intraclasse minimum, et non le critère du pourcentage de mal classés minimum.

2.8 Nous supposerons, quitte à changer le sens des axes, que α le centre de gravité de A est dans le premier quadrant de coordonnées, i. e. $\alpha_1 > 0$, $\alpha_2 > 0$. Nous désignerons par 1,2,3,4 les 4 côtés du carré A, et par M_1 , M_2 , M_3 , M_4 l'intersection de Δ avec les droites supports respectivement de 1,2,3,4 (cf figure 3).

La masse de A mal classée est égale au facteur $p/4a^2$ près à l'aire S de la portion de surface de A située sous Δ, cette portion de surface étant suivant les cas, soit vide, soit un triangle, soit un trapèze (cf figure 3).

Le tableau ci-dessous où $\rho^2 = (\alpha^1)^2 + (\alpha^2)^2$ désigne le carré de la distance de α à Δ, donne les coordonnées des M_i , avec les conditions pour que M_i soit sur le côté i de A :

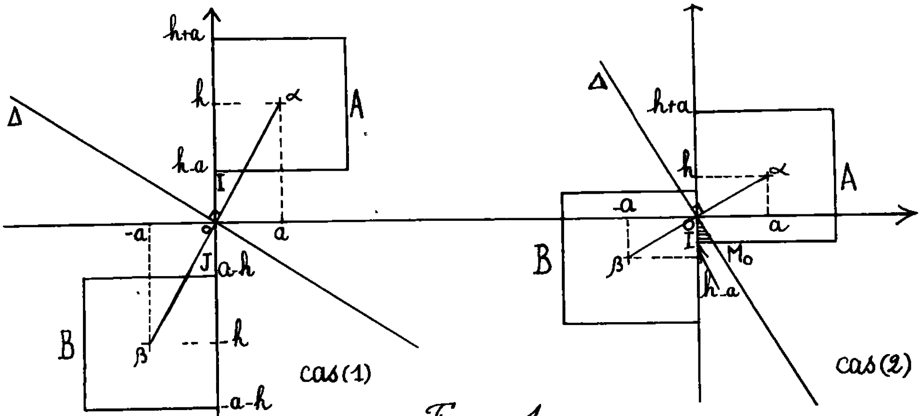

Figure 1

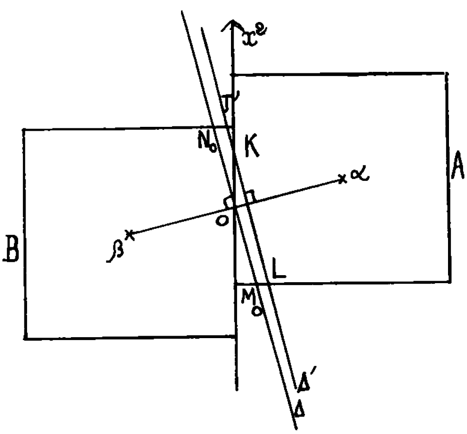

Figure 2: Toute droite Δ' parallèle à Δ discrimine moins bien A et B que Δ. par contre Ox^2 sépare parfaitement les 2 carrés, contrairement à Δ

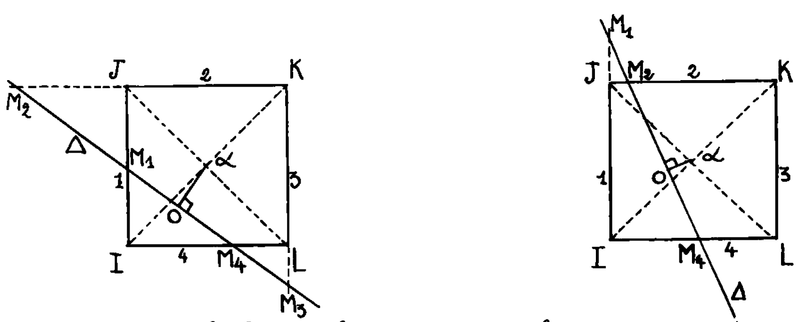

Figure 3: Calcul de la masse de la portion du carré A située de l'autre côté de α par rapport à Δ (portion de A mal classée) dans le cas général! On a représenté 2 cas de figure sur les 4 possibles

$$M^1 : (\alpha^1 - a \ ; \ -\alpha^1(\alpha^1 - a)/\alpha^2) \ ; \alpha^2 - a \le -\alpha^1(\alpha^1 - a)/\alpha^2 \le \alpha^2 + a$$
$$\text{ou } a(\alpha^1 - \alpha^2) \le \rho^2 \le a(\alpha^1 + \alpha^2)$$
$$M^2 : (-\alpha^2(\alpha^2 + a)/\alpha^1 \ ; \ \alpha^2 + a) \ ; \alpha^1 - a \le -\alpha^2(\alpha^2 + a)/\alpha^1 \le \alpha^1 + a$$
$$\text{ou } \rho^2 \le a(\alpha^1 - \alpha^2)$$
$$M^3 : (\alpha^1 + a \ ; \ -\alpha^1(\alpha^1 + a)/\alpha^2) \ ; \ \alpha^2 - a \le -\alpha^1(\alpha^1 + a)/\alpha^2 \le \alpha^2 + a$$
$$\text{ou } \rho^2 \le a(\alpha^2 - \alpha^1)$$
$$M^4 : (-\alpha^2(\alpha^2 - a)/\alpha^1 \ ; \ \alpha^2 - a) \ ; \alpha^1 - a \le -\alpha^2(\alpha^2 - a)/\alpha^1 \le \alpha^1 + a$$
$$\text{ou } a(\alpha^2 - \alpha^1) \le \rho^2 \le a(\alpha^1 + \alpha^2)$$

Il est immédiat de voir que Δ ne coupe jamais le carré A suivant $M_2 M_3$, ni suivant $M_1 M_2$ (sauf si $M_1 = M_2 = J$, auquel cas Δ coupe en fait A suivant $M_1 M_4 = M_2 M_4$), ni suivant $M_3 M_4$ (sauf si $M_3 = M_4 = L$, auquel cas Δ coupe en fait A suivant $M_1 M_3 = M_1 M_4$)

1° cas : Δ coupe A suivant $M_1 M_4$:

Ce cas est réalisé si : $a|\alpha^1 - \alpha^2| \le \rho^2 \le a(\alpha^1 + \alpha^2)$; (4)

on a alors pour S l'aire du triangle rectangle I $M_1 M_4$:

$$S = (-(\alpha^1(\alpha^1 - a)/\alpha^2) - (\alpha^2 - a))(-(\alpha^2(\alpha^2 - a)/\alpha^1) - (\alpha^1 - a))/2$$
$$= (a(\alpha^1 + \alpha^2) - \rho^2)^2/2\,\alpha^1\,\alpha^2 \qquad (5)$$

2° cas : Δ coupe A suivant $M_1 M_3$:

Ce cas est réalisé si : $\rho^2 \le a(\alpha^2 - \alpha^1)$; (6)

on a alors pour S l'aire du trapèze I $M_1 M_3$ L :

$$S = I\,L\,(M_1 I + M_3 L)/2 = a(-(\alpha^1(\alpha^1 - a + \alpha^1 + a)/\alpha^2) - 2(\alpha^2 - a))$$
$$= 2\,a(a\,\alpha^2 - \rho^2)/\alpha^2$$

3° cas : Δ coupe A suivant $M_2 M_4$:

Ce cas est réalisé si $\rho^2 \le a(\alpha^1 - \alpha^2)$ (7)

on a alors pour S l'aire du trapèze I J $M_2 M_4$:

$$S = I\,J(J\,M_2 + I\,M_4)/2 = a(-(\alpha^2(\alpha^2 + a + \alpha^2 - a)/\alpha^1) - 2(\alpha^1 - a))$$
$$= 2\,a(a\,\alpha^1 - \rho^2)/\alpha^1$$

4° cas : Δ ne coupe pas A :

Ce cas est réalisé si : $\rho^2 > a(\alpha^1 + \alpha^2)$ (8)

Dans ce cas S est nul.

Si $\alpha^1 = a$, $\alpha^2 = h$, (6) et (7) ne sont jamais vérifiées ; (4) est vérifié si $h \le a$, auquel cas d'après (5)

$S = (a(a+h) - a^2 - h^2)^2/2\,a\,h = h(h-a)^2/2\,a$, tandis que (8) est vérifié si $a < h$. On retrouve bien les résultats du § 2.5.

Pour calculer la masse de B mal classée, on procéderait de façon similaire.