Qüestió - V. 7, n.º 1 (març 1983) pp. 353-358

DIMENSIONALITAT EUCLIDIANA DE LES DISTÀNCIES ULTRAMÈTRIQUES

CARLES M. CUADRAS, FRANCESC CARMONA

UNIVERSITAT DE BARCELONA

Demostrem que tota distància ultramètrica definida en un conjunt de n elements, és representada en un espai euclidià amb dimensió n-1. Obtenim també alguns resultats sobre els valors propis de la matriu de productes escalars associada a la distància.

ULTRAMETRIC DISTANCE; EUCLIDEAN SPACE.

Key words: Multidimensional scaling:Mathematical taxonomy.

1. INTRODUCCIÓ.

Sigui E = {1,2,...n} un conjunt finit. Es diu que $U=(u_{ij})$ és una distància ultramètrica no degenerada sobre E, si u_{ij} representa una distància entre els elements i,j de E tal que

$$\text{a) } u_{ii} = 0 \quad \forall\, i \in E,$$

$$\text{b) } u_{ij} > 0 \quad \forall\, i \neq j,$$

$$\text{c) } u_{ij} \leq \max\{u_{ik}, u_{jk}\} \quad \forall\, i,j,k \in E. \qquad (1)$$

Es diu que $D = (d_{ij})$ és una distancia euclidiana m-dimensional sobre E si existeix una matriu de coordenades.

$$X = \begin{pmatrix} x_{11}\cdots x_{1m} \\ \cdots\cdots \\ x_{n1}\cdots x_{nm} \end{pmatrix}$$

on m és el mínim senser tal que

$$d_{ij}^2 = \sum_{h=1}^{m} (x_{ih}-x_{jh})^2 \qquad i,j=1,\ldots,n \qquad (2)$$

L'objectiu d'aquest treball és demostrar que tota distància ultramètrica no degenerada sobre un conjunt de n elements és euclidiana (n-1)-dimensional.

La motivació d'aquesta propietat té relació amb les tècniques multivariables que permeten, mitjançant una transformació lineal, ajustar unes coordenades X a unes coordenades Y, resultat de dos mètodes diferents de representació multidimensional de dades. El criteri d'ajust és el dels mínims cuadrats i la mesura de comparació és

$$\mu^2 = \text{tr}\,(X.X')+\text{tr}(Y.Y')-2\text{tr}(X'YY'X) \qquad (3)$$

Aquest criteri va ésser establert per Schönemann i Carroll /9/, Gower /4/, i ha estat utilitzat per Mardia /7/, Lingoes i Borg /6/.

Hom sap /1/ que un algorisme de classificació jeràrquica consisteix en modificar una distància inicial $D=(d_{ij})$ fins a obtenir una ultramètrica $U=(u_{ij})$, que aleshores queda representada per un dendrograma (fig. 1). Per tenir mida de la bona classificació obtinguda, el procediment clàssic consisteix en calcular la correlació cofenètica, és a dir, el coeficient de correlació entre els elements (d_{ij}) i els (u_{ij}). Com un altra mesura més objectiva, Gower /4/ proposà associar a la distància D una configuració euclidiana representada per unes coordenades X, associar a U un altre configuració euclidiana Y, i utilitzar aleshores la quantitat (3). Un aspecte essencial d'aquest procediment es que $U=(u_{ij})$ sigui realment una distància euclidiana, propietat que Gower /4/ pren com a una conjectura, però que no demostra. Aquest criteri ha estat con-

- Carles M. Cuadras i
- Francesc Carmona - Facultat de Biologia - Dep. de Bioestadística - Universitat de Barcelona - Av. Diagonal,637 Barcelona.

- Article rebut el Juny del 1983.

siderat per Rohlf /8/ com un dels mètodes -- per a comparar classificacions.

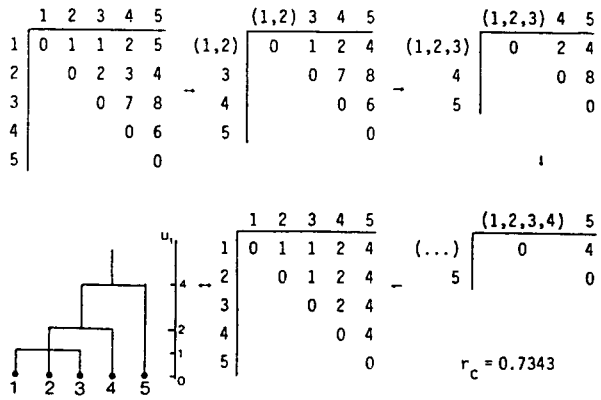

Fig. 1. Exemple de construcció d'una ultra-- mètrique per el mètode del mínim i dendrograma corresponent.

2. CARACTERITZACIO DE LA PROPIETAT EUCLIDIANA.

Per reconèixer si $D=(d_{ij})$ és una distància - euclidiana, cal aplicar el teorema 1. Siguin

$$\underline{1}=(1,\ldots,1)' \quad J=\underline{1}.\underline{1}' \quad H=I-\frac{1}{n}J \quad A=(a_{ij}) \qquad (4)$$

essent

$$a_{ij} = -\frac{1}{2} d_{ij}^2 ,$$

I = matriu identitat d'ordre n.

TEOREMA 1: Considerem la matriu B=HAH. Aleshores si B és semi-definida positiva de ---- ran(B) = m, D és euclidiana m-dimensional i les coordenades X verificant (2) són els elements d'una matriu nxm tal que

$$B = XX'$$

Reciprocament, si B té valors propis negatius, aleshores D es distància no euclidiana.

DEMOSTRACIO: Veure /1/.

En el cas que D és euclidiana, un procedi--- ment per obtenir les coordenades que verifiquen (2) consisteix en trobar els valors propis de B

$$\lambda_1 \geq \ldots \geq \lambda_m \geq \lambda_{m+1} = \ldots = \lambda_m = 0.$$

Aleshores els vectors propis λ- normalitzats formen les columnes d'una matriu X que ens - dona la configuració euclidiana buscada. --- Aquest procediment s'anomena "mètode de les coordenades principals".

3. DIMENSIONALITAT DE LES ULTRAMETRIQUES -- EUCLIDIANES.

Sigui $U=(u_{ij})$ una ultramètrica i f una fun-- ció monòtona creixent. Aleshores

$$u^*_{ij} = f(u_{ij})$$

verifica

$$u^*_{ij} \leq u^*_{i'j'} \text{ si i només si } u_{ij} \leq u_{i'j'} \qquad (5)$$

i la propietat (1) segueix essent certa. Direm doncs que dues ultramètriques són equivalents si existeix una funció f verificant -- (5). Indiquem $[u_{ij}]$ la classe de les ultramè triques que són equivalents a u_{ij}. Es evi--- dent que les ultramètriques d'una mateixa -- classe d'equivalència defineixen la mateixa classificació jeràrquica, variant només en - l'índex de la jerarquia.

LEMA. Existeix un representant de $[u_{ij}]$ que és distància euclidiana.

DEMOSTRACIO: És conseqüència de que sem-- pre és possible transformar una distància segons una funció monòtona per convertir-- la en euclidiana. Dos exemples de transfor macions són:

a) Lineal:

$$u^*_{ij} = \alpha u_{ij} + \beta \qquad /3/$$

b) Additiva:

$$(u^*_{ij})^2 = u^2_{ij} - 2\alpha \qquad /5/ \text{ i } /7/$$

Tenim doncs que u_{ij} és euclidiana, o bé -- equivalent a alguna ultramètrica euclidiana. Ara bé, provar que tot altre representant de $[u_{ij}]$ es distància euclidiana, el qual vol dir que qualsevol ultramètrica té la propietat euclidiana, no és tan evident. Els teoremes 2 i 3 ens demostren aquesta - propietat.

TEOREMA 2: Tota ultramètrica euclidiana defi nida sobre un conjunt E de n elements és --- (n-1)-dimensional.

DEMOSTRACIÓ: Per hipótesi, existeixen n punts $P_1,\ldots,P_n$ d'un espai euclidià tal - que

$$u_{ij} = d(P_i,P_j)$$

on d significa distància euclidiana. Considerem ara els vectors $V_1,\ldots,V_{n-1}$, essent

$$V_i = \overrightarrow{P_1P_{i+1}} \qquad i=1,\ldots,n-1.$$

Hem de demostrar que $V_1,\ldots,V_{n-1}$ són vec-- tors linealment independents. Ho farem pro vant que la matriu de productes escalars associada a $\langle V_i\rangle$ és definida positiva.

Suposem que u_{12} és la mínima distància entre els elements de E

$$u_{12} = \text{mín}\{u_{ij} \mid i \neq j, \quad i,j \in E\}$$

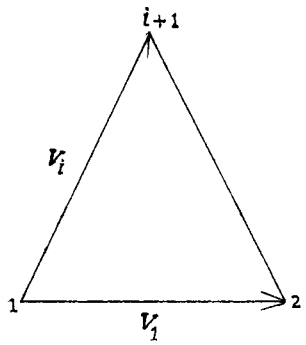

Fig. 2: El triangle (1,2, i 1) es isòsceles com a conseqüencia de la propietat ultramètrica.

Aleshores (1) implica

$$u_{12} \leq u_{1i+1} = u_{2i+1}$$

i per tant tenim (fig. 2) que $V_i - V_1 = \overrightarrow{P_2P_{i+1}}$ verifica

$$\| V_i - V_1 \| = \| V_i \|$$

i aleshores, indicant $V_h.V_k$ el producte escalar euclidià, tenim

$$V_i.V_i = V_i.V_i + V_1.V_1 - 2V_i.V_1$$

doncs

$$V_1.V_i = \frac{1}{2} V_1.V_1 = \frac{u_{12}^2}{2} = a \qquad i=2,\ldots,n-1$$

A més a més, de

$$\| V_i - V_j \|^2 = \| V_i \|^2 + \| V_j \|^2 - 2V_i.V_j$$

deduim

$$V_i.V_j = \frac{1}{2}(u_{1i+1}^2 + u_{1j+1}^2 - u_{i+1j+1}^2) \qquad 1<i\neq j$$

L'expressió de la matriu simètrica $\sum$, -- d'ordre (n-1)x(n-1), de productes escalars, és

$$\sum = \begin{pmatrix} 2a & a & a & a & \\ & u_{13}^2 & (u_{13}^2+u_{14}^2-u_{34}^2)/2 & (u_{13}^2+u_{15}^2-u_{35}^2)/2 & \cdots \\ & & u_{14}^2 & (u_{14}^2+u_{15}^2-u_{45}^2)/2 & \cdots \\ & & & \cdots\cdots\cdots & \end{pmatrix} \tag{6}$$

Seguidament demostrarem per inducció que - el més petit valor propi λ_1 de $\sum$ verifica

$$\lambda_1 \geq \frac{u_{12}^2}{2} = a \tag{7}$$

Sigui

$$P(\lambda) = \det(\sum - \lambda I)$$

el polinomi característic. Per a n=3 és

$$P(\lambda) = \lambda^2 - (2a+u_{13}^2)\lambda + (2au_{13}^2 - a^2),$$

i la seva derivada verifica:

$$P'(\lambda) = -(u_{13}^2-\lambda) - (u_{12}^2-\lambda) < 0 \text{ si } \lambda \leq a.$$

A més

$$P(a) = a(u_{13}^2 - u_{12}^2) \geq 0.$$

Això ens demostra que $P(\lambda)$ és decreixent - per tot $\lambda \leq a$, i no negativa per $\lambda=a$. Per -- tant $P(\lambda)$ s'anul.larà només per valors $\lambda \geq a$.

Suposem ara que (7) és certa per n-1. De-- mostrem que llavors ho serà també per n. - La derivada del polinomi característic és

$$P'(\lambda) = -P_1(\lambda) - \ldots - P_{n-1}(\lambda)$$

on $P_j(\lambda)$ és el polinomi característic de - la matriu que s'obté suprimint la fila j i la columna j de $\sum$. Per hipótesi d'induc-- ció, el menor valor propi d'aquesta matriu verifica $\lambda_1 \geq a$ i per tant

$$P_j(\lambda) \geq 0 \quad \text{si} \quad \lambda \leq a \quad j=1,\ldots,n-1 \tag{8}$$

Això és clar per $j>1$. Per $j=1$, al suprimir la primera fila i columna de $\sum$, no obtenim una estructura similar a (6), però mitjançant un canvi de base apropiat, que deixarà invariant els valors propis, podem -- afirmar que $P_1(\lambda)\geq 0$ si $\lambda \leq u_{hk}^2/2$, essent -- $u_{hk}=\min\{u_{ij} \mid 2\neq i\neq j\neq 2\}\geq a$. Així (8) és vàlid també per $j=1$.

Podem afirmar llavors que

$$P'(\lambda)\leq 0 \quad \text{si} \quad \lambda \leq a \qquad (9)$$

que vol dir que $P(\lambda)$ és decreixent per $\lambda \leq a$.

Passem a calcular $P(a)$. Restant a la diagonal de $\sum$ el número a i tot seguit restant a cada fila la primera, ens queda un determinant que podem desenvolupar per la primera columna.

$$\Delta = \begin{vmatrix} a & a & a & a & \\ 0 & (u_{13}^2-u_{12}^2) & (u_{13}^2+u_{14}^2-u_{34}^2-u_{12}^2)/2 & (u_{13}^2+u_{15}^2-u_{35}^2-u_{12}^2)/2 & \cdots \\ 0 & & (u_{14}^2-u_{12}^2) & (u_{14}^2+u_{15}^2-u_{45}^2-u_{12}^2)/2 & \cdots \\ \vdots & \cdots & & & \end{vmatrix}$$

$$\Delta = a \det(\bar{\sum})$$

Considerem ara la ultramètrica $\bar{u}_{ij}$ tal que

$$\bar{u}_{ij}^2 = u_{ij}^2 - u_{12}^2 \qquad 2\neq i\neq j\neq 2$$
$$= 0 \qquad i=j\neq 2.$$

Podem veure que $\bar{\sum}$ és la matriu de productes escalars de n-2 vectors associats a la ultramètrica $\bar{u}_{ij}$, i que té la mateixa estructura que $\sum$, però d'ordre (n-2). En efecte, si suposem que

$$u_{13} = \min\{u_{ij} \mid 2\neq i\neq j\neq 2\}$$

tindrem que

$$\sum = \begin{pmatrix} \bar{u}_{13}^2 & \bar{u}_{13/2}^2 & \bar{u}_{13/2}^2 & \cdots \\ & \bar{u}_{14}^2 & (\bar{u}_{14}^2+\bar{u}_{15}^2-\bar{u}_{45}^2)/2 & \cdots \\ & \cdots & & \end{pmatrix}$$

En cas que aquest mínim fos $u_{hk}(h\neq 1)$, efectuant un canvi de base apropiat, arribaríem a la mateixa conclusió. Si $\bar{u}_{ij}$ és degenerada ($\bar{u}_{ij}=0$ per a algún $i\neq j$) es tindrà ----- $\det(\bar{\sum}) = 0$. En cas contrari $\det(\bar{\sum})>0$. Tenim doncs

$$P(a) = a.\det(\bar{\sum})\geq 0$$

Finalment, com que a més a més $P(\lambda)$ és decreixent per $\lambda\leq a$, tenim que $P(\lambda)\geq 0$ per a tot $\lambda\leq a$. Això demostra que $\lambda_1\geq a$, $\sum$ és definida positiva ran$(\sum)=n-1$. En conseqüencia, la dimensió euclidiana de u_{ij} és (n-1).

4. TOTA ULTRAMÈTRICA ÉS EUCLIDIANA.

La demostració que tota ultramètrica u_{ij} és euclidiana és una conseqüència del teorema 3.

TEOREMA 3: Sigui u_{ij} una distància ultramètrica. Tot representant de la classe $[u_{ij}]$ - distància euclidiana.

DEMOSTRACIO: Ho demostrarem per reducció a l'absurd. Suposem que u_{ij} no és euclidiana. Considerem el representant de $[u_{ij}]$

$$\bar{u}_{ij}^2 = u_{ij}^2-2\alpha>0 \quad \text{si} \quad i\neq j,$$
$$= 0 \qquad \text{si} \quad i=j.$$

on α és una constant. Amb les notacions -- del teorema 1, siguin

$$A=(a_{ij}) \quad \bar{A}=(\bar{a}_{ij}) \quad a_{ij}=-\frac{1}{2}u_{ij}^2 \quad \bar{a}_{ij}=-\frac{1}{2}\bar{u}_{ij}^2$$

Aleshores

$$\bar{A} = A-\alpha(I-J)$$

i també és fàcil veure que

$$J = \frac{1}{n}J.J \quad J.H = 0 \quad H.H = H$$

i per tant

$$\bar{B} = H\bar{A}H = (I-\frac{1}{n}J)(A-\alpha(I-J))(I-\frac{1}{n}J) = B-\alpha H$$

Com que hem suposat u_{ij} no euclidiana, els valors propis i vectors propis de B=HAH serien

$$\lambda_1\geq\ldots\geq\lambda_r>0\geq\lambda_1'\geq\ldots\geq\lambda_s'$$

$$U_1,\ldots,U_r,\underline{1},V_1,\ldots,V_s$$

$\underline{1}$ és vector propi de valor propi 0, aleshores

$$\underline{1}'.U_i = \underline{1}'.V_i = 0$$

i per tant

$$\bar{B}U_i = BU_i - \alpha(I-\frac{1}{n}J)U_i = (\lambda_i-\alpha)U_i \quad i=1,\ldots,r$$

$$\bar{B}V_j = (\lambda_j-\alpha)V_j \qquad j=1,\ldots,s$$

$$\bar{B}\underline{1} = 0$$

Els valors i vectors propis de $\bar{B}$ són

$$\lambda_1-\alpha \geq \ldots \geq \lambda_r-\alpha \geq \lambda'_1-\alpha \geq \ldots \geq \lambda'_s-\alpha,\ 0$$

$$U_1,\ldots,U_r\ ;\ V_1,\ldots,V_s\ ;\ \underline{1}$$

Prenem $\alpha=\lambda'_s$. Aleshores $\bar{B}$ serà semidefinida positiva i tendrá com a mínim 2 valors propis nuls. La matriu $\bar{B}$ está lligada a la distància ultramètrica no degenerada

$$\bar{u}_{ij}^2 = u_{ij}^2 + 2|\lambda'_s| > 0 \quad \text{si} \quad i\neq j,$$

$$= 0 \qquad\qquad \text{si} \quad i=j,$$

que, d'acord amb el teorema 1, serà euclidiana m-dimensional, on $m\leq n-2$. Però això ens contradiu el teorema 2. Hem d'admetre doncs que u_{ij}, i tot representant de $[u_{ij}]$, és distància euclidiana.

Finalment, podem obtenir un resultat relatiu al mínim valor propi no nul de la matriu de productes escalars B associada a una ultramètrica u_{ij}.

TEOREMA 4: Sigui u_{ij} una distància ultramètrica i sigui $B = HAH$ la matriu introduida en el teorema 1. Es verifica:

1) B es semidefinida positiva,

2) ran(B) = n-1,

3) El menor valor propi no nul de B és

$$\lambda_{n-1} = \frac{\bar{u}^2}{2} \qquad (10)$$

essent $\bar{u} = \min\{u_{ij} \mid i\neq j\}$.

DEMOSTRACIO: Els apartats 1) i 2) son conseqüencia dels teoremes anteriors. Suposem que (10) no sigui certa. Donat $\varepsilon>0$, considerem la ultramètrica equivalent

$$\bar{u}_{ij}^2 = u_{ij}^2 - \bar{u}^2 + \varepsilon > 0 \qquad i\neq j,$$

$$= 0 \qquad\qquad i=j.$$

Seguint els mateixos pasos del teorema anterior agafant

$$\alpha = \frac{\bar{u}^2}{2} - \frac{\varepsilon}{2},$$

els valors propis no nuls de la matriu $\bar{B}$ associada a $\bar{u}_{ij}$ seran

$$\bar{\lambda}_1 = \lambda_1 - \frac{\bar{u}^2}{2} + \frac{\varepsilon}{2} \geq \ldots \geq \bar{\lambda}_{n-1} = \lambda_{n-1} - \frac{\bar{u}^2}{2} + \frac{\varepsilon}{2}$$

Aleshores, si fos $\lambda_{n-1} < \bar{u}^2/2$, podem trobar un ε suficientment petit per a que sigui $\bar{\lambda}_{n-1}<0$, i això entraria en contradicció amb el teorema 3. Observem també que per $\varepsilon=0$ s'obté una ultramètrica $\bar{u}_{ij}$ degenerada, i per tant $\bar{\lambda}_{n-1}=0$ que implica (10).

5. CONCLUSIONS.

Hom sap que una distància mètrica, es a dir, verificant la desigualtat triangular, no és necessàriament euclidiana. En aquest treball hem demostrat que si es verifica la propietat més restrictiva (1), anomenada desigualtat ultramètrica, aleshores la distància es sempre euclidiana. Això només es vàlid quan está definida sobre un conjunt finit. Aquesta propietat té conseqüencies dins les tècniques de comparació de mètodes d'ordenació i classificació de dades.

Seguint un camí diferent, /2/ obté l'estructura dels valors i vectors propis de la matriu B de productes escalars associada a una ultramètrica U sobre un conjunt E.Troba també que les coordenades principals de U donen representacions espacials dels elements de E, que descriuen geomètricament l'estructura jerárquica sobre E definida per U.

6. BIBLIOGRAFIA.

/1/ CUADRAS, C.M., "Métodos de Análisis Multivariante", Eunibar,Barcelona, 642pp., -- (1981).

/2/ CUADRAS, C.M., "Análisis algebraico sobre distancias ultramétricas", Com. pres. sl 44º Per de Ses. del Inst.Intern. de Estadística, Madrid, (1983).

/3/ CUADRAS, C.M., RUIZ-RIVAS, C. "Una contribución al análisis de proximidades" - P. Secc. Mat. U. Aut. Barcelona, 22, 103-106, (1980).

/4/ GOWER, J.C. "Statistical methods of comparing different multivariate analysis - of the same data", A: Mathematics in the Archaelogical and Historical Sciences, - (Hodson, F.R., Kendall, D.G. i Tautu, P., eds.), Edinburgh University Press, Edinburgh, pp. 138-149, (1971).

/5/ LINGOES, J.C. "Some boundary conditions for a monotone analysis of symmetric matrices", Psychometrika, 36, 195-203, --- (1971).

/6/ LINGOES, J.C., BORG, I. "A direct ------ approach differences scaling using in--- creasingly complex transformations" ---- Psychometrika, 43(4), 491-519, (1978).

/7/ MARDIA, K.V., "Some properties of classical multidimensional scaling", Comm.Stat. A7(13), 1233-1241, (1978).

/8/ ROHLF, F.J.,"Methods of comparing classifications", Ann. Rew. Ecol. System., 5, 101-113, (1974).

/9/ SCHONEMANN, P.H., CARROLL, R.M., "Fitting one matrix to another under choice of a - central dilation and a rigid motion" ---- Psychometrika, 35, 245-255, (1970).