QÜESTIIÓ, Vol 19, 1, 2, 3 pp. 131-151, 1995

PIRÁMIDES INDEXADAS Y DISIMILARIDADES PIRAMIDALES

CARLES CAPDEVILA MARQUÈS* y ANTONI ARCAS PONS‡

En este trabajo se pretende una formalización de las bases matemáticas sobre las que se amparan los modelos de clasificación y representación piramidal, estableciéndose para ello una equivalencia entre los conceptos de pirámide indexada y disimilaridad piramidal. Asimismo se precisa el concepto de información redundante, contenida en una estructura piramidal y se dan criterios objetivos para poder prescindir de ella sin que ello suponga una modificación de la estructura piramidal.

Indexed Pyramids and Pyramidal Dissimilarities.

Key words: Pyramid, Pyramidal dissimilarity, Spare groups.

1. INTRODUCCIÓN

Los modelos piramidales de clasificación y representación de datos multidimensionales, introducidos por Bertrand y Diday (1985) y estudiados más recientemente por otros autores, (Fichet, Durand, etc.) pretenden ser una generalización de los modelos jerárquicos clásicos, en el sentido de permitir la existencia no sólo de grupos disjuntos o encajados, sino también la existencia de grupos solapados y por tanto permitir clasificaciones en las que, a cada nivel, los grupos en que queda dividida la población no tengan que constituir forzosamente particiones de la misma, como ocurre en el caso de las clasificaciones inducidas por jerarquías, sino que puedan

* Carles Capdevila Marquès. Departament de Matemàtica. Universitat de Lleida. C/. Bisbe Messeguer, s/n. 25003. Lleida.

‡ Antoni Arcas Pons. Departament d'Estadística. Universitat de Barcelona. Avda. Diagona, 645. 08028 Barcelona.

–Article rebut el juny de 1994.

–Acceptat el maig de 1995.

constituir recubrimientos. En este caso puede suceder que un mismo individuo, o grupo de individuos, pertenezcan a dos grupos diferentes de una misma clasificación, con lo cual estos individuos podrán caracterizarse por las propiedades de los diversos grupos a los que pertenezcan. En este sentido podemos intuir que estos nuevos modelos de clasificación y representación de datos, deberán ser más próximos a la realidad que los modelos jerárquicos clásicos.

En este trabajo, pretendemos dar una visión rigurosa, desde el punto de vista de la formalización matemática, de las bases teóricas sobre las que se amparan los modelos piramidales de clasificación y representación. Más concretamente, se formaliza la relación existente entre pirámide indexada y disimilaridad piramidal.

Desde otro punto de vista se plantea también el problema de la información redundante contenida en una estructura piramidal y que, como tal, puede ser eliminada de la estructura simplificándola. En este sentido se precisa el concepto de información redundante y se dan criterios objetivos para prescindir de ella sin tener que modificar la pirámide correspondiente.

2. PRELIMINARES

En esta sección introducimos los conceptos básicos para el estudio de los modelos piramidales de clasificación y representación de datos, cuales son los de disimilaridad piramidal, pirámide, pirámide indexada, preorden compatible, etc. así como algunas propiedades de los mismos.

Definición 2.1

Dada una disimilaridad d sobre un conjunto finito Ω, diremos que es una disimilaridad piramidal, si verifica las condiciones:

D.1 $d(\omega_i,\omega_j) \geq 0, \quad \forall\, \omega_i,\omega_j \in \Omega$

D.2 $d(\omega_i,\omega_i) = 0, \quad \forall\, \omega_i \in \Omega$

D.3 $d(\omega_i,\omega_j) = d(\omega_j,\omega_i), \quad \forall\, \omega_i,\omega_j \in \Omega$

D.4 Existe un preorden total $(\leq)$ sobre Ω, tal que para todo $\omega_i,\omega_j,\omega_k \in \Omega$, con $\omega_i \leq \omega_j \leq \omega_k$, $d(\omega_i,\omega_k) \geq \max\{d(\omega_i,\omega_j),\ d(\omega_j,\omega_k)\}$

Proposición 2.1

Una matriz simétrica de orden n, de términos reales positivos (d_{ij}), es una matriz asociada a una disimilaridad ultramétrica d, sobre un conjunto $\Omega = \{\omega_1,\ldots,\omega_n\}$, sii *existe una ordenación total de los individuos* ($\omega_1 \leq \cdots \leq \omega_n$), *de modo que la matriz* $M(d,\leq)$[1] *verifique:*

a) $d_{ij} \leq d_{ij+1}$ *para todo* $i \leq j$.

b) *Para todo* $k \in \{1,\ldots,n\}$, *si* $d_{kk+1} = \ldots = d_{kk+s+1} < d_{kk+s+2}$, *entonces:*
$d_{k+1j} \leq d_{kj}$ *para* $k+1 < j \leq k+s+1$
$d_{k+1j} = d_{kj}$ *para* $j > k+s+1$

La demostración puede encontrarse en Lerman (1981).

Proposición 2.2

Toda disimilaridad ultramétrica es también piramidal.

Demostración

Si d es una disimilaridad ultramétrica sobre un conjunto finito Ω, podemos considerar la ordenación sobre Ω establecida por la proposición 2.1.

Sean $\omega_i, \omega_j, \omega_k \in \Omega$ tales que $\omega_i \leq \omega_j \leq \omega_k$. Las condiciones a) y b) de la proposición anterior implican, respectivamente: $d(\omega_i,\omega_j) \leq d(\omega_i,\omega_k)$ y $d(\omega_j,\omega_k) \leq d(\omega_i,\omega_k)$. Por tanto, de ambas desigualdades obtendremos: $d(\omega_i,\omega_k) \geq \max\{d(\omega_i,\omega_j), d(\omega_j,\omega_k)\}$, con lo cual d será piramidal.

■

Definición 2.2

Dada una disimilaridad d, y un preorden $\leq$, definidos sobre Ω, diremos que d y $\leq$ son compatibles, si y solamente si, para cualesquiera elementos de Ω, tales que $\omega_i \leq \omega_j \leq \omega_k$, se verifica, $d(\omega_i,\omega_k) \geq \max\{d(\omega_i,\omega_j), d(\omega_j,\omega_k)\}$.

Definición 2.3

Sea Ω un conjunto finito preordenado y $h \in \wp(\Omega)$. Diremos que h es una parte conexa de Ω respecto al preorden, sii $\forall\, \omega_i, \omega_j \in h$, con $\omega_i \leq \omega_j$, $\{\omega \in \Omega / \omega_i \leq \omega \leq \omega_j\} \subset h$.

Definición 2.4

Diremos que un preorden definido sobre Ω es compatible con una familia $\mathrm{H} \subset \wp(\Omega)$, sii $\forall\, h \in \mathrm{H}$, h es una parte conexa respecto al preorden.

[1]Si d es una disimilaridad sobre un conjunto finito Ω, cualquier ordenación total ($\leq$) de sus elementos, dará lugar a una matriz de disimilaridad para d, en la que la i-ésima columna (fila) representará al i-ésimo elemento de Ω respecto a la ordenación dada. Esta matriz se representa por $\mathrm{M}(d,\leq)$.

Proposición 2.3

Una disimilaridad d es una disimilaridad piramidal sobre Ω, sii existe un preorden total sobre Ω compatible con d.

La proposición es evidente a partir de las definiciones.

Definición 2.5

Sea Ω un conjunto finito y $\mathbf{P} \subset \wp(\Omega)$. Diremos que $\mathbf{P}$ es una pirámide, si se verifican las condiciones siguientes:

P.1 $\Omega \in \mathbf{P}$.
P.2 $\forall\, \omega \in \Omega, \quad \{\omega\} \in \mathbf{P}$.
P.3 $\forall\, h,h' \in \mathbf{P}, \quad h \cap h' = \emptyset$ ó $h \cap h' \in \mathbf{P}$.
P.4 Existe un preorden total sobre Ω, compatible con $\mathbf{P}$.

Definición 2.6

Sea $\mathbf{P}$ una pirámide sobre Ω.

Diremos que una función $i : \mathbf{P} \to \mathbb{R}^+$ es un índice asociado a $\mathbf{P}$, si verifica las condiciones:

I.1 $i(\{\omega\}) = 0, \quad \forall\, \omega \in \Omega$
I.2 $\forall\, h,h' \in \mathbf{P}, \quad h \subset h' \quad \Rightarrow \quad i(h) \geq i(h')$

Si $\mathbf{P}$ es una pirámide sobre Ω e i un índice asociado a $\mathbf{P}$, diremos que $(\mathbf{P},i)$ es una pirámide indexada.

Definición 2.7

Si $(\mathbf{P},i)$ es una pirámide indexada tal que:

I.3 $\forall\, h,h' \in \mathbf{P}, \quad h \subsetneq h' \quad \Rightarrow \quad i(h) < i(h')$

diremos que $(\mathbf{P},i)$ es una pirámide indexada en sentido estricto.

Si en lugar de la condición I.3 se verifica:

I.3′ $\forall\, h,h' \in \mathbf{P}$ tales que $h \subsetneq h'$ y $i(h) = i(h')$,
existen $h_1,h_2 \in \mathbf{P}$, distintos de h, y tales que $h = h_1 \cap h_2$.

Entonces diremos que $(\mathbf{P},i)$ es una pirámide indexada en sentido amplio.

De las definiciones anteriores resulta inmediato:

Proposición 2.4

Toda pirámide indexada en sentido estricto, lo es también en sentido amplio.

A los elementos de una pirámide los llamaremos "*grupos*" de la pirámide, y si una pirámide es indexada, el valor del índice sobre cada grupo, se denomina "*índice del grupo*".

Es interesante observar, tal como se demuestra en Diday (1986), que toda jerarquía total indexada es también una pirámide indexada, con lo cual puede concluirse que las pirámides son una generalización natural de las jerarquías.

En este sentido, puede verse que cada grupo de una pirámide puede tener hasta dos predecesores, mientras que en una jerarquía cada clase tiene un único predecesor.

Del mismo modo que los dendrogramas proporcionan una representación visual de las Jerarquías Indexadas, es posible construir otro tipo de grafos conexos, a los que llamaremos "*grafos piramidales*" o "*pirámides*", que proporcionen una representación visual de las Pirámides Indexadas.

Su construcción consiste en expresar gráficamente las propiedades características de las pirámides, es decir:

Dos grupos, no encajados, pueden tener intersección no vacía.
Cada grupo puede tener hasta dos predecesores.
Existencia de un preorden total para el que cada grupo sea conexo.

El grafo piramidal expresará entonces, en un solo diagrama, la imbricación de unos grupos en otros y el índice de la pirámide. Su representación visual se efectuará asociando a cada grupo un segmento horizontal situado, sobre una escala, a una altura igual a la de su índice y unido a sus predecesores mediante unas líneas oblicuas, que llamaremos "*aristas*", de manera que cada arista una el centro del segmento asociado a un grupo con uno de los extremos de los segmentos asociados a sus predecesores.

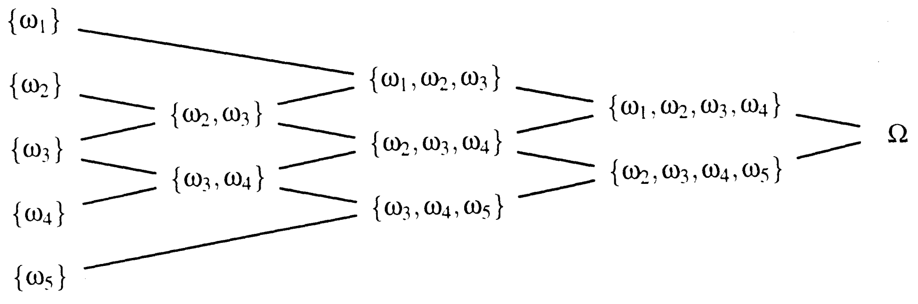
Ejemplo 2.1

Sea $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$, ordenado por $\omega_1 \leq \omega_2 \leq \omega_3 \leq \omega_4 \leq \omega_5$ y sea

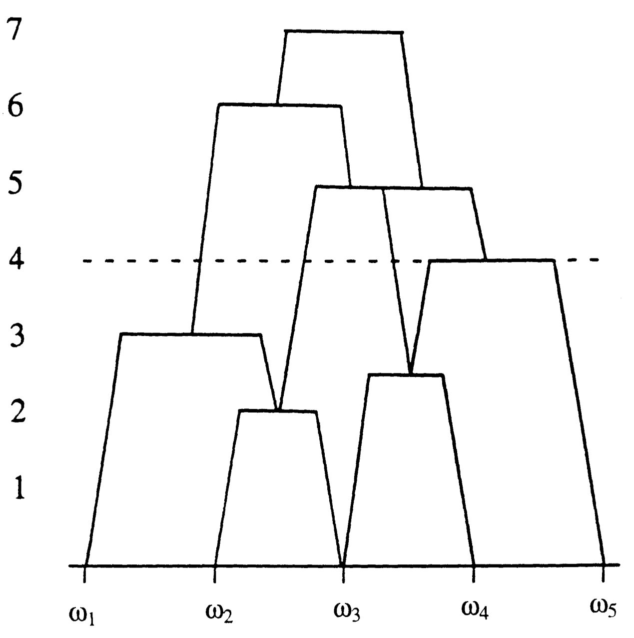
$\mathbf{P} = \{\{\omega_1\}, \ldots, \{\omega_5\}, \{\omega_2, \omega_3\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_3, \omega_4, \omega_5\}, \{\omega_2, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_4, \omega_5\}, \{\omega_1, \omega_2, \omega_3, \omega_4\}, \Omega\}$ una pirámide indexada por:

$$
\begin{array}{ll}
i(\{\omega_j\}) = 0, \quad \text{para} \quad j \in \{1, \ldots, 5\} & i(\{\omega_2, \omega_3, \omega_4\}) = 5 \\
i(\{\omega_2, \omega_3\}) = 2 & i(\{\omega_2, \omega_3, \omega_4, \omega_5\}) = 5 \\
i(\{\omega_3, \omega_4\}) = 2.5 & i(\{\omega_1, \omega_2, \omega_3, \omega_4\}) = 6 \\
i(\{\omega_1, \omega_2, \omega_3\}) = 3 & i(\Omega) = 7 \\
i(\{\omega_3, \omega_4, \omega_5\}) = 4 &
\end{array}
$$

Si partimos de los singletones, $\{\omega_1\},\ldots,\{\omega_5\}$ y determinamos los sucesivos predecesores, obtendremos:



La representación visual de la pirámide $(\mathbf{P},i)$ será:



Obsérvese como, a cada nivel, la clasificación correspondiente constituye un recubrimiento de la población Ω. A nivel 4, por ejemplo, la clasificación piramidal obtenida viene dada por los grupos: $\{\omega_1,\omega_2,\omega_3\},\{\omega_2,\omega_3\},\{\omega_3,\omega_4\},\{\omega_3,\omega_4,\omega_5\}$, que obviamente forman un recubrimiento de Ω.

3. RELACIÓN ENTRE PIRÁMIDES INDEXADAS Y DISIMILARIDADES PIRAMIDALES

3.1. Planteamiento del problema

Teniendo en cuenta que la base matemática de las representaciones jerárquicas se fundamenta en la relación existente entre la obtención de una disimilaridad ultramétrica sobre una población Ω y una jerarquía indexada de $\wp(\Omega)$, en esta sección demostraremos la existencia de una correspondencia biunívoca, entre las pirámides indexadas y las disimilaridades piramidales, de modo que los grafos piramidales puedan interpretarse como la representación visual de dichas disimilaridades.

Para ello, tendremos en cuenta el paralelismo existente con el caso de las jerarquías y las ultramétricas antes mencionado.

Sea Ω un conjunto finito y d una disimilaridad piramidal sobre Ω. Nuestro objetivo será el de asociar a la disimilaridad d una pirámide indexada de $\wp(\Omega)$.

Sea $A = \{\alpha \in \mathbb{R}/\exists\ \omega_i, \omega_j \in \Omega \quad \text{y} \quad d(\omega_i, \omega_j) = \alpha\}$, para cualquier $\alpha \in A$, definimos la siguiente relación sobre Ω:

$$\forall\, \omega, \omega' \in \Omega; \qquad \omega R_\alpha \omega' \quad \Leftrightarrow \quad d(\omega, \omega') \leq \alpha$$

Si d fuese una disimilaridad ultramétrica, $\forall\, \alpha \in A$ las relaciones R_α serían de equivalencia. En nuestro caso, si d es piramidal, las propiedades reflexiva y simétrica se satisfacen de forma inmediata, sin embargo ello no es suficiente para asegurar la transitividad de dichas relaciones, por lo que no serán de equivalencia, y por tanto no darán lugar, en general, a particiones de Ω, sino a ciertos recubrimientos, a los que denominaremos P_α. Así pues, $\forall\, \alpha \in A$ el recubrimiento P_α estará formado por grupos $h \in \wp(\Omega)$ tales que $d(\omega, \omega') \leq \alpha, \quad \forall\, \omega, \omega' \in h$.

A pesar de todo, podríamos pensar que la unión de todos estos recubrimientos, $P' = \bigcup_{\alpha \in A} P_\alpha$ es la pirámide inducida por la disimilaridad d. Pues bien, esto no siempre es cierto, puesto que la intersección de dos grupos de un tal P', no siempre es de P' o vacía, tal como puede verse en el ejemplo 3.1.

Ejemplo 3.1

Sea $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ y d una disimilaridad piramidal sobre Ω, dada por la matriz:

$$\begin{bmatrix} 0 & 1 & 1 & 3 \\ & 0 & 1 & 2 \\ & & 0 & 2 \\ & & & 0 \end{bmatrix}$$

En este caso, $A = \{0,1,2,3\}$ y construyendo los distintos recubrimientos de Ω, para los distintos valores de α, tendremos:

$$\begin{aligned} P_0 &= \{\{\omega_1\},\{\omega_2\},\{\omega_3\},\{\omega_4\}\} \\ P_1 &= \{\{\omega_1,\omega_2,\omega_3\},\{\omega_4\}\} \\ P_2 &= \{\{\omega_1,\omega_2,\omega_3\},\{\omega_2,\omega_3,\omega_4\}\} \\ P_3 &= \{\Omega\} \end{aligned}$$

con lo que $P' = \{\{\omega_1\},\{\omega_2\},\{\omega_3\},\{\omega_4\},\{\omega_1,\omega_2,\omega_3\},\{\omega_2,\omega_3,\omega_4\},\Omega\}$ y puesto que existe una intersección que no es de P', $\omega_1,\omega_2,\omega_3\} \cap \{\omega_2,\omega_3,\omega_4 = \{\omega_2,\omega_3\} \notin P'$, P' no es una pirámide.

Así pues deberemos reconsiderar el método de construcción de una pirámide a partir de una disimilaridad piramidal.

En este empeño, para cualquier $\alpha \in A$, consideraremos las siguientes familias de $\wp(\Omega)$:

$$\begin{aligned} \mathbf{C} &= \{h \in \wp(\Omega)/\exists\, \omega_i,\omega_j \in \Omega; \quad d(\omega_i,\omega_j) = \alpha, \quad h = \{\omega \in \Omega/\omega_i \leq \omega \leq \omega_j\}\}^2 \\ \mathbf{C} &= \{h \in \mathbf{C}_\alpha/h \text{es maximal en } \mathbf{C}_\alpha, \text{ en el sentido de la inclusión}\} \end{aligned}$$

Precisando un poco más, $h \in \mathbf{C}_\alpha$ sii existen $\omega_i,\omega_j \in \Omega$, a distancia α, y $h = \{\omega \in \Omega/\omega_i \leq \omega \leq \omega_j\} = [\omega_i,\omega_j]_\alpha$. Por consiguiente, $h \in \mathbf{C}^*_\alpha$ sii $\exists\, \omega_i,\omega_j \in \Omega/d(\alpha_i,\alpha_j) = \alpha$ y $h = [\omega_i,\omega_j]_\alpha$ maximal. Es decir, los grupos $h \in \mathbf{C}^*_\alpha$ son intervalos maximales, en el sentido de que $\forall\, \omega' \notin h, \quad \exists \omega \in h$ tal que $d(\omega,\omega') > \alpha$.

Obsérvese que en el caso particular de la población Ω y de la disimilaridad piramidal d del ejemplo 3.1, estas familias serían:

$$\begin{aligned} \mathbf{C}_0 &= \{\{\omega_1\},\{\omega_2\},\{\omega_3\},\{\omega_4\}\} & \mathbf{C} &= \{\{\omega_1\},\{\omega_2\},\{\omega_3\},\{\omega_4\}\} \\ \mathbf{C}_1 &= \{\{\omega_1,\omega_2\},\{\omega_1,\omega_2,\omega_3\},\{\omega_{2,3}\}\} & \mathbf{C} &= \{\{\omega_1,\omega_2,\omega_3\}\} \\ \mathbf{C}_2 &= \{\{\omega_2,\omega_3,\omega_4\},\{\omega_3,\omega_4\}\} & \mathbf{C} &= \{\{\omega_2,\omega_3,\omega_4\}\} \\ \mathbf{C}_3 &= \{\Omega\} & \mathbf{C} &= \{\Omega\} \end{aligned}$$

A partir de estas nuevas familias de $\wp(\Omega)$, podríamos intentar definir la pirámide inducida por d como $P^* = \bigcup_{\alpha \in A} \mathbf{C}^*_\alpha$; pero P^*, en general, tampoco será una pirámide puesto que en el caso particular de la población y la disimilaridad piramidal considerados en el ejemplo 3.1, resulta ser $P^* = P'$, con lo cual P^* tampoco será la pirámide buscada.

---

[2] Esta definición tiene sentido puesto que al ser d una disimilaridad piramidal, debe existir sobre Ω un preorden total, lo cual permite, dados dos individuos $\omega_i,\omega_j \in \Omega$, a distancia α, considerar los individuos $\omega \in \Omega$ tales que $\omega_i \leq \omega \leq \omega_j$.

Por otra parte, el resultado $P^* = P'$, que hemos obtenido en el caso particular del ejemplo 3.1, puede generalizarse, como se demuestra en la siguiente proposición.

Proposición 3.1.1

Si d es una disimilaridad piramidal sobre Ω, $A = \{\alpha \in \mathbb{R} / \exists\ \omega_i, \omega_j \in \Omega$ y $d(\omega_i, \omega_j) = \alpha\}$ y $\forall\, \alpha \in A$, $\mathbf{C}^*_\alpha$ y P_α, las familias de $\wp(\Omega)$ definidas anteriormente, entonces fijado $\alpha \in A$, se tiene $\mathbf{C}^*_\alpha = P_\alpha$.

Demostración

Veamos en primer lugar que fijado $\alpha \in \mathrm{A}, \mathbf{C}^*_\alpha \subset \mathrm{P}_\alpha$.

Sea $h \in \mathbf{C}^*_\alpha$, esto significa que existirán $\omega_i, \omega_j \in \Omega$ a distancia α, tales que $h = [\omega_i, \omega_j]_\alpha$ maximal, y por tanto $\forall\, \omega' \notin h$, $\exists\, \omega \in h$ con $d(\omega, \omega') > \alpha$, lo cual equivale a decir que $\forall\, \omega \in h$ que esté a distancia menor o igual que α de cualquier otro $\omega' \in \Omega$, deberá ser $\omega' \in h$, con lo cual tendremos que $h = \{$individuos de Ω con interdistancias $\leq \alpha\}$ y por tanto será $h \in \mathrm{P}_\alpha$.

Si por el contrario partimos de un grupo $h \in \mathrm{P}_\alpha$, será $h = \{$individuos de Ω con interdistancias $\leq \alpha\}$. Puesto que $\alpha \in \mathrm{A}$, existirán $\omega_i, \omega_j \in \Omega$ tales que $d(\omega_i, \omega_j) = \alpha$ y supongamos que ω_i y ω_j sean los individuos de Ω extremos, (en términos del preorden total existente en Ω, asociado a d) a distancia α. En estas condiciones, $\forall\, \omega \in h$ deberá ser $\omega_i \leq \omega \leq \omega_j$, puesto que si fuese $\omega < \omega_i \leq \omega_j$, por ser d piramidal tendríamos $d(\omega, \omega_j) \geq \alpha$. Por otra parte, puesto que $d(\omega_i, \omega_j) = \alpha$, ω_i y ω_j serán de h, y puesto que ω también es de h, será $d(\omega, \omega_j) \leq \alpha$. Por tanto tendríamos $d(\omega, \omega_j) = \alpha$ y $\omega < \omega_i \leq \omega_j$, con lo cual ω_i y ω_j no serían los individuos extremos a distancia α.

Así pues, $h \subset \{x \in \Omega / \omega_i \leq x \leq \omega_j\} = [\omega_i, \omega_j]_\alpha$.

Por otra parte, $\forall\, x \in \Omega$ tal que $\omega_i \leq x \leq \omega_j$ con $d(\omega_i, \omega_j) = \alpha$, deberá ser $x \in h$, puesto que $\forall\, x' \in [\omega_i, \omega_j]_\alpha$, $d(x, x') \leq d(\omega_i, \omega_j)$ por ser d piramidal, y por tanto $d(x, x') \leq \alpha$, con lo cual tendremos que $x \in h$.

De ambas inclusiones resultará $h = [\omega_i, \omega_j]_\alpha$. Además, al ser ω_i y ω_j los individuos extremos a distancia α, el intervalo $[\omega_i, \omega_j]_\alpha$ será maximal en $\mathbf{C}_\alpha$ respecto a la inclusión.

Así pues, $h = [\omega_i, \omega_j]_\alpha$ maximal $\Rightarrow$ $h \in \mathbf{C}^*_\alpha$.

Si hubiésemos supuesto $\omega_i \leq \omega_j < \omega$, habríamos llegado a la misma conclusión.

■

3.2. Teoremas de Existencia y Unicidad

Teorema 3.2.1

Toda disimilaridad piramidal d sobre Ω, define una pirámide indexada en sentido amplio de $\wp(\Omega)$ que es única.

Demostración

Dada una disimilaridad piramidal d sobre Ω, definimos el siguiente subconjunto de $\wp(\Omega)$: $\mathbf{P} = \mathrm{P}^* \cup \{$intersecciones no vacías de dos grupos de $\mathrm{P}^*\}$

Con ello, tendremos que los conjuntos de $\mathbf{P}$ o bien son grupos de P^*, o bien son intersecciones de dos grupos de P^*, que son precisamente los que le faltan a éste para ser una pirámide.

Veamos, en primer lugar, que $\mathbf{P}$ es una pirámide.

P.1 $\Omega \in \mathbf{P}$.
En efecto, $\Omega \in \mathrm{P}^*$ puesto que será $\Omega \in \mathbf{C}^*_\alpha$, con $\alpha = \max(\mathrm{A})$.

P.2 $\forall\, \omega \in \Omega; \quad \{\omega\} \in \mathbf{P}$
$\forall\, \omega \in \Omega, \quad \{\omega\} \in \mathrm{P}^*$ puesto que $\{\omega\} \in \mathbf{C}^*_\alpha$ y $0 \in \mathrm{A}$.

P.3 $\forall\, h, h' \in \mathbf{P}, \quad h \cap h' = \emptyset$ ó $h \cap h' \in \mathbf{P}$.
Sean h, h' dos grupos de $\mathbf{P}$ tales que $h \cap h' \neq \emptyset$
Puesto que cualquier grupo de $\mathbf{P}$, o bien es de P^*, o es intersección no vacía de dos grupos de P^*, si h y h' son grupos de $\mathbf{P}$, existirán $h_i \in \mathrm{P}^*$, $i \in \{1,2,3,4\}$, tales que: $h = h_1 \cap h_2$ y $h' = h_3 \cap h_4$.
Puesto que d es piramidal, tiene sentido suponer la existencia de un preorden total sobre Ω. Consideremos, para cualquier $i \in \{1,2,3,4\}$, ω_i y ω'_i los individuos extremos de $h_i(\omega_i \leq \omega'_i)$.
Si consideramos $\quad j = \max\{\omega_i\}_{i=1,2,3,4}$
$\omega'_k = \min\{\omega'_i\}_{i=1,2,3,4}$
resulta inmediato, a partir de la consideración de que el preorden sobre es total, que: $h \cap h' = h_j \cap h_k$, con $h_j, h_k \in \mathrm{P}^*$, (Ver Fig. 3.2.1).
Así pues, tendremos que $h \cap h' \in \mathbf{P}$.
$h \cap h' = (h_1 \cap h_2) \cap (h_3 \cap h_4) = h_1 \cap h_4$:

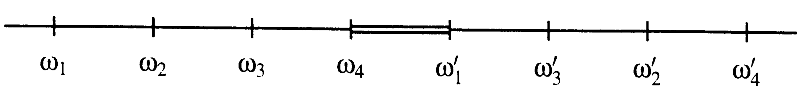


Figura 3.2.1

P.4 Existe un preorden total sobre Ω, compatible con $\mathbf{P}$.
Puesto que d es piramidal, existe un preorden total sobre Ω compatible con d, y este mismo preorden es compatible con $\mathbf{P}$, puesto que cualquier grupo de P^* es conexo respecto a este preorden por construcción, y la intersección no vacía de dos grupos conexos es otro grupo conexo (puesto que el preorden es total). Por tanto cualquier grupo de $\mathbf{P}$ será conexo respecto al preorden.
Finalmente vamos a ver que esta pirámide puede ser indexada en sentido amplio. Para ello, definamos la siguiente aplicación:

$i: \mathbf{P} \longrightarrow \mathbb{R}^+$

$\quad h \longrightarrow i(h) = \max\{d(\omega,\omega'); \quad \omega,\omega' \in h\}$

Entonces se satisface:

I.1 $\forall \in \Omega, \quad i(\{\omega\}) = 0.$
Inmediato a partir de la definición.

I.2 $\forall h, h' \in \mathbf{P}, \quad$ si $h \subset h', \quad$ entonces $i(h) \leq i(h')$.
Si $h \subset h'$ es evidente que $\max\{d(x,y); \quad x,y \in h\} = i(h)$ es menor o igual que $\max\{d(x,y); x,y \in h'\} = i(h')$, con lo cual $i(h) \leq i(h')$.

I.3 $\forall h, h' \in \mathbf{P}$ con $h \subsetneq h'$ e $i(h) = i(h')$, existen $h_1, h_2 \in \mathbf{P}$, distintos de h, y tales que $h = h_1 \cap h_2$.
Si $h \subsetneq h'$ y son del mismo índice, entonces $h \notin P^*$, puesto que h no es maximal para $\alpha = i(h)$.
Por tanto, si $h \in \mathbf{P}$ y $h \notin P^*$, existirán h_1 y h_2 de $P^* \subset \mathbf{P}$, distintos de h, y tales que $h = h_1 \cap h_2$.

Así pues, $(\mathbf{P}, i)$ es una pirámide indexada en sentido amplio que además es única por construcción.

■

Teorema 3.2.2

Toda pirámide indexada $(\mathbf{P}, i)$ de $\wp(\Omega)$, induce una única disimilaridad piramidal sobre Ω.

Demostración

Dada una pirámide indexada $(\mathbf{P}, i)$, para cada par de individuos ω y ω' de Ω, definimos: $d(\omega,\omega') = i(h_0)$, siendo h_0, el mínimo grupo de $\mathbf{P}$, en el sentido de la inclusión, que contiene ω y ω'.

Esta definición es correcta puesto que si existe un único grupo de $\mathbf{P}$ que contiene ω y ω', entonces éste es el mínimo, y si existen varios, su intersección, que también es de **P**, es el mínimo, con lo cual tiene sentido considerar el mínimo grupo de **P** que contiene ω y ω'.

Veamos pues que d es una disimilaridad piramidal:

D.1 $\forall\, \omega, \omega' \in \Omega, \quad d(\omega,\omega') \geq 0.$
Inmediato, puesto que para cada $h \in \mathbf{P}, i(h) \geq 0$.

D.2 $\forall\, \omega, \omega' \in \Omega, \quad d(\omega,\omega) = 0.$
El menor grupo de **P** que contiene ω, es $\{\omega\}$, e $i(\{\omega\}) = 0$.

D.3 $\forall\, \omega, \omega' \in \Omega, \quad d(\omega,\omega') = d(\omega',\omega).$
Inmediato.

D.4 Existe un preorden total sobre Ω tal que, dados $\omega, \omega', \omega''$ de Ω, con $\omega \leq \omega' \leq \omega''$, $d(\omega,\omega'') \geq \max\{d(\omega,\omega'), d(\omega',\omega'')\}$.
Puesto que **P** es una pirámide, existe en Ω un preorden total compatible con **P**. Sean ω, ω' y ω'' elementos de Ω, con $\omega \leq \omega' \leq \omega''$.

Sea $h_{\omega\omega''}$ el mínimo grupo de **P** que contiene ω y ω'', puesto que $h_{\omega\omega''}$ es conexo respecto al preorden que **P** induce sobre Ω, tendremos que $\omega' \in h_{\omega\omega''}$.

Si $h_{\omega\omega'}$ es el mínimo grupo de **P** que contiene ω y ω', tendremos que tanto $h_{\omega\omega''}$ como $h_{\omega\omega'}$ contienen ω y ω'. Si son grupos conexos y $h_{\omega\omega'}$ es el mínimo, tendremos que $h_{\omega\omega'} \subset h_{\omega\omega''}$, por tanto $i(h_{\omega\omega'}) \leq i(h_{\omega\omega''})$ y $d(\omega,\omega') \leq d(\omega,\omega'')$.

Por otra parte, $h_{\omega\omega''}$ y $h_{\omega'\omega''}$, ambos contienen a ω' y ω'', por tanto, $h_{\omega'\omega''} \subset h_{\omega\omega''}$, así pues $i(h_{\omega'\omega''}) \leq i(h_{\omega\omega''})$ y $d(\omega',\omega'') \leq d(\omega,\omega'')$.

De ambas desigualdades, obtenemos: $d(\omega,\omega'') \geq \max\{d(\omega,\omega'), d(\omega',\omega'')\}$, que es lo que queríamos demostrar.

■

Puesto que, dados dos individuos cualesquiera ω y ω' de Ω, el mínimo grupo de **P** que los contiene es único, la disimilaridad d definida como el índice de dicho grupo, será también única.

Finalmente, cabe resaltar que el teorema anterior es válido, en particular, para pirámides indexadas en sentido amplio. Así pues, los teoremas 3.2.1 y 3.2.2 establecen una correspondencia biunívoca entre el conjunto de las pirámides indexadas en sentido amplio y el de las disimilaridades piramidales.

3.3. Otras Relaciones

Proposición 3.3.1

Sea $(\mathbf{P}, i)$ *una pirámide indexada de* $\wp(\Omega)$ *y sean* ω *y* ω' *dos individuos cualesquiera de* Ω. *El mínimo grupo de* **P** *que contiene a* ω *y* ω', *es también el de menor índice de entre todos los que los contienen.*

Proposición 3.3.2

Sea h un grupo cualquiera de una pirámide **P**. *Si ω y ω' son los individuos extremos de h, entonces h es el mínimo grupo de* **P** *que los contiene.*

Las demostraciones de las proposiciones 3.3.1 y 3.3.2 son inmediatas a partir de la definición de pirámide indexada y del hecho que cada grupo de la pirámide sea conexo respecto un cierto preorden total definido sobre la población Ω.

Proposición 3.3.3

Sea $(\mathbf{P}, i)$ una pirámide indexada en sentido amplio sobre Ω, y d la disimilaridad piramidal inducida. Sea $A = \{\alpha \in \mathbb{R} / \exists\ \omega_i, \omega_j \in \Omega \quad y \quad d(\omega_i, \omega_j) = \alpha\}$ y sea P' la familia de $\wp(\Omega)$ definida a partir de la disimilaridad d en la sección 3.1. En estas condiciones:

$$h \in \mathbf{P} \quad \Leftrightarrow \quad \left[\begin{array}{l} h \in P' \\ o \\ \exists\, h_1, h_2 \in P', \quad \textit{distintos de } h, \textit{ tales que } h = h_1 \cap h_2 \end{array} \right.$$

Demostración

Sea h un grupo cualquiera de **P**. Puesto que **P** es una pirámide, tiene sentido considerar los individuos extremos de h respecto al preorden total correspondiente definido sobre Ω, sean éstos ω_i y ω_j y supongamos que $d(\omega_i, \omega_j) = \alpha_0$. En estas condiciones, la proposición 3.3.2 nos asegura que h es el mínimo grupo de **P** que contiene ω_i y ω_j.

Supongamos que $h \notin$ P', entonces $\forall\, \alpha \in$ A, $h \notin \mathrm{P}_\alpha$, en particular, $h \notin \mathrm{P}_{\alpha 0}$, y por tanto, existirá $\omega_0 \in \Omega$, tal que:

$$(1) \qquad \omega_0 \notin h \quad \text{y} \quad \forall\, x \in h, \quad d(\omega_0, x) \leq \alpha_0 = d(\omega_i, \omega_j)$$

Si ω_i, ω_j son los individuos más alejados de h, y $\omega_0 \notin h$, entonces no puede ser $\omega_i \leq \omega_0 \leq \omega_j$, y por tanto tendrá que ser: $\omega_0 < \omega_i \leq \omega_j$ o bien $\omega_i \leq \omega_j < 0$.

Supongamos $\omega_0 < \omega_i \leq \omega_j$. Puesto que d es piramidal, tendremos: $d(\omega_0, \omega_j) \geq d(\omega_i, \omega_j)$.

Por otra parte, para $x = \omega_j$ la desigualdad (1) quedará: $d(\omega_0, \omega_j) \leq d(\omega_i, \omega_j)$.

Y combinando ambas desigualdades, tendremos:

$$(2) \qquad d(\omega_0, \omega_j) = d(\omega_i, \omega_j)$$

Consideremos, además, h_0, el mínimo grupo de **P** que contiene ω_0 y ω_j.

Puesto que la disimilaridad piramidal d, viene inducida por la pirámide $\mathbf{P}$, resultará que: $d(\omega_0,\omega_j) = i(h_0)$ y $d(\omega_i,\omega_j) = i(h)$.

De lo cual, y teniendo en cuenta (2), se deduce que $i(h) = i(h_0)$.

Veamos que $h \subsetneq h_0$:

Puesto que $\omega_0 < \omega_i \leq \omega_j$ y h_0 es conexo, ω_i también ha de pertenecer a h_0.

Así pues, h es el mínimo grupo que contiene ω_i y ω_j y h_0 también los contiene, por tanto $h \subset h_0$.

Además $\omega_0 \in h_0$ y $\omega_0 \notin h$, con lo cual, $h \subsetneq h_0$.

Resumiendo pues, tendremos que: dado $h \in \mathbf{P}$, si $h \notin P'$, existirá $h_0 \in \mathbf{P}$ tal que $h \subsetneq h_0$ e $i(h) = i(h_0)$, y puesto que $(\mathbf{P}, i)$ es una pirámide indexada en sentido amplio, existirán $h_1, h_2 \in \mathbf{P}$, distintos de h, tales que $h = h_1 \cap h_2$.

Ahora bien, nuestro objetivo era demostrar que si $h \notin P'$, entonces es intersección de dos grupos de P', y no de $\mathbf{P}$.

Supongamos pues que por lo menos uno de los dos grupos, h_1 por ejemplo, no es de P'. Entonces, existirán $h_1', h_1'' \in \mathbf{P}$, distintos de h_1 y tales que $h_1 = h_1' \cap h_1''$, con lo cual $h = h_1' \cap h_1'' \cap h_2$.

Si, a su vez, alguno de estos grupos no es de P', h_1' por ejemplo, existirán $\overline{h_1'}$ y $\hat{h_1'}$ de $\mathbf{P}$, distintos de h_1', con $h_1' = \overline{h_1'} \cap \hat{h_1'}$, con lo cual $h = \overline{h_1'} \cap \hat{h_1'} \cap h_1'' \cap h_2$, y así sucesivamente.

En este proceso, obsérvese que $h \subsetneq h_1 \subsetneq h_1' \subsetneq \overline{h_1'} \subsetneq \cdots$ Puesto que las inclusiones son estrictas y $\wp(\Omega)$ es finito, o bien llegaremos a un grupo de P', o bien a un grupo $\tilde{h}_1$ tal que $h \subsetneq h_1 \subsetneq \cdots \tilde{h}_1$ y no podrá desdoblarse como intersección de dos grupos de $\mathbf{P}$, entonces tendrá que ser $\tilde{h}_1 \in P'$, puesto que de lo contrario, sería intersección de dos grupos estrictamente más grandes con lo cual $\tilde{h}_1$ no sería el último grupo de la cadena como estamos suponiendo.

Así pues, aplicando este proceso a los grupos que no sean de P', tendremos que h puede expresarse como $h = \tilde{h}_1 \cap \tilde{h}_2 \cap \cdots \cap \tilde{h}_r$, con $\tilde{h}_i \in P'$ y $\tilde{h}_1 \neq h$, para todo $i \in \{1,2,\ldots,r\}$.

A partir de aquí, por ser el preorden sobre Ω total, es inmediato demostrar que $h = \tilde{h}_k \cap \tilde{h}_j$, con $k,j \in \{1,\ldots,r\}$, y $\tilde{h}_k, \tilde{h}_j \in P'$ que es lo que queríamos demostrar.

■

Sea $h \in P'$, entonces $\exists\, \alpha \in A / h \in P_\alpha$ y por tanto será $h = \{$individuos de Ω con interdistancias $\leq \alpha\}$.

Si $\alpha \in A \quad \Rightarrow \quad \exists\, \omega_i, \omega_j \in \Omega$ tales que $d(\omega_i, \omega_j) = \alpha$ y supongamos que ω_i y ω_j son los individuos extremos de h, en términos del preorden total sobre Ω asociado a d, puesto que si no fuese así, es decir si existiese un $\overline{\omega} \in h$ tal que $\overline{\omega} < \omega_i \leq \omega_j$ (ó bien $\omega_i \leq \omega_j < \overline{\omega}$) tendríamos que, por ser d piramidal, $d(\overline{\omega}, \omega_j) \geq d(\omega_i, \omega_j)$; y por ser $\overline{\omega}, \omega_j \in h$, $d(\overline{\omega}, \omega_j) \leq d(\omega_i, \omega_j)$, con lo cual $d(\overline{\omega}, \omega_j) = d(\omega_i, \omega_j)$ y por tanto, en la definición de h podríamos sustituir ω_i por $\overline{\omega}$.

Supongamos pues, ω_i, ω_j los extremos de h y sea h_0 el mínimo grupo de **P** que los contiene.

Entonces, $\forall\, \omega \in h$ será $\omega_i \leq \omega \leq \omega_j$ y puesto que h_0 es conexo, $\omega \in h_0$, con lo cual, $h \subset h_0$.

Si d es la disimilaridad piramidal inducida por la pirámide $(\mathbf{P}, i)$, la inclusión anterior no podrá ser estricta, por tanto $h = h_0$ y en consecuencia, $h \in \mathbf{P}$.

Finalmente, si $h = h_1 \cap h_2$, con $h_i \in P'$ y $h_i \neq h$, para $i \in \{1,2\}$, puesto que acabamos de demostrar que $P' \subset \mathbf{P}$, resultará que h será intersección de dos grupos de **P**, por tanto será $h \in \mathbf{P}$.

Teorema 3.3.1

La pirámide indexada en sentido amplio inducida por una disimilaridad piramidal, induce esa misma disimilaridad piramidal.

Demostración

Para ello, vamos a ver que si una pirámide indexada en sentido amplio, $(\mathbf{P}, i)$ induce una disimilaridad piramidal d', y a su vez es la pirámide inducida por otra disimilaridad piramidal d, entonces $d = d'$.

Si d' es la disimilaridad piramidal inducida por $(\mathbf{P}, i)$, dados dos individuos cualesquiera $x, y \in \Omega$, $\quad d'(x,y) = i(h_{xy})$, siendo h_{xy} el mínimo grupo de **P** que contiene a x e y.

Si $(\mathbf{P}, i)$ es la pirámide inducida por d, entonces $i(h_{xy}) = \max\{d(\omega, \omega'); \omega, \omega' \in h_{xy}\}$, por consiguiente será $i(h_{xy}) \geq d(x,y)$.

De donde, $d'(x,y) \geq d(x,y)$.

Vamos a ver que esta desigualdad no puede ser estricta.

Supongamos que existan $x_0, y_0 \in \Omega$ tales que $d'(x_0,y_0) > d(x_0,y_0)$. Si $d(x_0,y_0) = \alpha$ podemos considerar la familia de $\wp(\Omega)$, $\mathbf{C}^*_\alpha$ tal como la hemos definido en la sección 3.1. Sean $h^1_\alpha, \dots, h^r_\alpha$ los grupos de dicha familia, que a su vez serán todos ellos grupos de $\mathbf{P}$, por ser ésta la pirámide inducida por d.

Por otra parte, es claro que existirá algún h^i_α, con $i \in \{1,\dots,r\}$ tal que $x_0, y_0 \in h^i_\alpha$.

Así pues, si $(\mathbf{P}, i)$ es la pirámide inducida por d, $\quad i(h^i_\alpha) = \alpha < d'(x_0,y_0) = i(h_0)$. Ahora bien, si h_0 es el mínimo grupo de $\mathbf{P}$ que contiene x_0 e y_0, en virtud de la proposición 3.3.1 es también el grupo de menor índice de entre todos los que contienen x_0 e y_0, por tanto no puede ser que $i(h^i_\alpha) < i(h_0)$. Por tanto no puede existir ningún par de individuos, $x_0, y_0 \in \Omega$ tales que $d'(x_0,y_0) > d(x_0,y_0)$, con lo cual tendrá que ser $d'(x,y) = d(x,y)$ para cualquier par, x e y, de individuos de Ω, y por tanto $d = d'$.

■

Teorema 3.3.2

La disimilaridad piramidal inducida por una pirámide indexada en sentido amplio, induce esa misma pirámide.

Demostración

Para ello, veremos que si la pirámide $(\mathbf{P}, i)$ induce la disimilaridad piramidal d, y a su vez, ésta induce la pirámide $(\tilde{\mathbf{P}}, \tilde{i})$, entonces $\mathbf{P} = \tilde{\mathbf{P}}$ y $i = \tilde{i}$.

Si $(\tilde{\mathbf{P}}, \tilde{i})$ es la pirámide indexada inducida por d, por la demostración del teorema 3.2.1 sabemos que:

$$h \in \tilde{\mathbf{P}} \quad \Leftrightarrow \quad \begin{bmatrix} h \in \tilde{\mathrm{P}^*} \\ \mathrm{o} \\ \exists\, h_1, h_2 \in \tilde{\mathrm{P}^*}, \quad \text{distintos de } h, \text{ tales que } h = h_1 \cap h_2 \end{bmatrix}$$

Por otra parte, puesto que gracias a la proposición 3.1.1, $\tilde{\mathrm{P}^*} = \mathrm{P}'$, resultará:

$$h \in \tilde{\mathbf{P}} \quad \Leftrightarrow \quad \begin{bmatrix} h \in \mathrm{P}' \\ \mathrm{o} \\ \exists\, h_1, h_2 \in \mathrm{P}', \quad \text{distintos de } h, \text{ tales que } h = h_1 \cap h_2 \end{bmatrix}$$

Y por la proposición 3.3.3 esto equivale a que $h \in \mathbf{P}$.

Para ver que $i = \tilde{i}$, tendremos en cuenta que $\mathbf{P} = \tilde{\mathbf{P}}$ y demostraremos que $\forall\, h \in \mathbf{P}$, $i(h) = \tilde{i}(h)$.

Si $(\tilde{\mathbf{P}}, \tilde{i})$ es la pirámide indexada inducida por d, para cualquier $h \in \mathbf{P} = \tilde{\mathbf{P}}$, $\tilde{i}(h) = \max\{d(\omega,\omega'),\ \forall\, \omega, \omega' \in h\}$, por tanto existirán $x, y \in h$ tales que $\tilde{i}(h) = d(x,y)$.

Por otra parte, sean ω_i y ω_j los individuos extremos de h. Puesto que $h \in \mathbf{P}$ y $\mathbf{P}$ es la pirámide que induce d, si ω_i y ω_j son los individuos más alejados de h, éste será el mínimo grupo de $\mathbf{P}$ que los contendrá (Proposición 3.3.2), y por tanto $i(h) = d(\omega_i, \omega_j)$.

Así pues, hasta el momento tenemos que: $x, y \in h \in \mathbf{P}$; ω_i, ω_j son los extremos de h; y d es la disimilaridad piramidal inducida por $(\mathbf{P}, i)$; con lo cual $d(x,y) \leq d(\omega_i, \omega_j)$.

Finalmente, puesto que $d(x,y) = \max\{d(\omega, \omega'), \quad \forall\ \omega, \omega' \in h\}$, tendremos que $d(x,y) \geq d(\omega_i, \omega_j)$.

De ambas desigualdades resultará: $d(x,y) = d(\omega_i, \omega_j)$ y por lo tanto $\forall\ h \in \mathbf{P}$, $i(h) = \bar{i}(h)$.

■

4. GRUPOS SOBRANTES. DEPURACIÓN DE UNA PIRÁMIDE

Dada una pirámide indexada $(\mathbf{P}, i)$ construida por alguno de los algoritmos de clasificación piramidal existentes, Diday (1986), Bertrand (1986), Capdevila (1993), podemos preguntarnos si existen grupos "sobrantes", es decir, grupos que proporcionen información redundante sobre la clasificación y por tanto que puedan ser eliminados de la pirámide sin que varíe la clasificación piramidal.

Para responder a la pregunta con el máximo rigor, será necesario, en primer lugar, precisar el concepto de "*grupo sobrante*" en una pirámide indexada.

Si $(\mathbf{P}, i)$ es una pirámide indexada en sentido amplio, es evidente que ningún grupo puede ser sobrante, puesto que a dicha pirámide le corresponde biunívocamente una disimilaridad piramidal d, por tanto si eliminamos algún grupo h de $\mathbf{P}$, a pesar de que $\mathbf{P} - \{h\}$ continuase siendo una pirámide indexada en sentido amplio, induciría otra disimilaridad piramidal d' distinta a d y por tanto una clasificación también distinta. Si por el contrario, la pirámide es indexada en sentido estricto, también lo será en sentido amplio (Proposición 2.3) y por consiguiente tampoco habrá grupos susceptibles de ser eliminados.

Sea pues $(\mathbf{P}, i)$ una pirámide indexada pero no en sentido amplio. En este caso, existirán grupos $h, h' \in \mathbf{P}$; $\quad h \subsetneq h'; i(h) = i(h')$ y h no será intersección no trivial de grupos de $\mathbf{P}$.

Consideremos el conjunto: $\quad \Sigma_{\mathrm{P}} = \{h \in \mathbf{P} / \exists\ h' \in \mathbf{P}; h \subsetneq h'; \quad i(h) = i(h')$ y h no es intersección no trivial de grupos de $\mathbf{P}\}$ que será no vacío, por ser la pirámide indexada no en sentido amplio.

En estas condiciones, es evidente que $\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}}$ será también una pirámide indexada y además, en sentido amplio.

Si finalmente demostramos que la disimilaridad piramidal inducida por $(\mathbf{P},i)$ es la misma que la inducida por $(\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}},i)$, quedará demostrado que los grupos de $\boldsymbol{\Sigma}_{\mathrm{P}}$ son realmente los grupos sobrantes de la pirámide $\mathbf{P}$, puesto que $\mathbf{P}$ y $\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}}$ serán pirámides equivalentes, en el sentido de tener asociada la misma disimilaridad piramidal, y al ser $(\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}},i)$ indexada en sentido amplio, no tendrá ningún grupo sobrante, con lo cual podremos asegurar que los grupos de $\boldsymbol{\Sigma}_{\mathrm{P}}$ serán los únicos grupos sobrantes de $\mathbf{P}$.

Teorema 4.1

La pirámide indexada sobre $\Omega(\mathbf{P},i)$, induce la misma disimilaridad piramidal que la pirámide $(\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}},i)$.

Demostración

Sabemos que cualquier pirámide indexada de $\wp(\Omega)$, sea en sentido amplio o no, induce una disimilaridad piramidal sobre Ω, sin más que definir la distancia entre dos individuos como el índice del menor grupo que los contiene.

Sea pues Ω un conjunto finito y $(\mathbf{P},i)$ una pirámide indexada. Si lo es en sentido amplio, entonces $\boldsymbol{\Sigma}_{\mathrm{P}}=\emptyset$ y no hay nada que demostrar.

Supongamos pues que $(\mathbf{P},i)$ es indexada pero no en sentido amplio, con lo cual $\boldsymbol{\Sigma}_{\mathrm{P}}\neq\emptyset$ y $\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}}\subsetneq\mathbf{P}$.

Sean d y d' las disimilaridades piramidales inducidas respectivamente por $(\mathbf{P},i)$ y $(\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}},i)$.

Sean x e y dos individuos cualesquiera de Ω, y sea h_0 el mínimo grupo de $\mathbf{P}$ que los contiene. Entonces será $d(x,y)=i(h_0)$.

Si $h_0\notin\boldsymbol{\Sigma}_{\mathrm{P}}$, entonces $h_0\in\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}}$, y por tanto $d'(x,y)=i(h_0)$, con lo cual tendremos $d(x,y)=d'(x,y)$, que es lo que queríamos ver.

Si por el contrario, $h_0\in\boldsymbol{\Sigma}_{\mathrm{P}},h_0\notin\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}}$, y además existirán otros grupos de $\mathbf{P}$ que contendrán estrictamente h_0 y de su mismo índice. Estos otros grupos lo serán, naturalmente, en número finito, y no todos ellos podrán ser de $\boldsymbol{\Sigma}_{\mathrm{P}}$ (por lo menos $\Omega\notin\boldsymbol{\Sigma}_{\mathrm{P}}$).

Sea pues, h'_0 el menor de los grupos de $\mathbf{P}$ tales que $h_0\subsetneq h'_0$, $i(h_0)=i(h'_0)$ y $h'_0\notin\boldsymbol{\Sigma}_{\mathrm{P}}$.

En estas condiciones pues, h'_0 será el menor grupo de $\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}}$ que contendrá h_0 y por tanto el menor grupo de $\mathbf{P}-\boldsymbol{\Sigma}_{\mathrm{P}}$ que contendrá x e y. Así pues, $d'(x,y)=i(h'_0)=i(h_0)$ y por tanto $d(x,y)=d'(x,y)$ que es lo que queríamos demostrar.

Resumiendo pues, la caracterización de los grupos sobrantes de una pirámide indexada será la siguiente: $h \in (\mathbf{P}, i)$ es un *grupo sobrante* sii $h \in \Sigma_{\mathrm{P}}$. Es decir, h es un *grupo sobrante*, si no es intersección no trivial de grupos de $\mathbf{P}$ y existe otro grupo $h' \in (\mathbf{P}, i)$ tal que: $h \subsetneq h'$ e $i(h) = i(h')$.

La Depuración de una Pirámide consistirá en eliminar los grupos sobrantes. Una vez eliminados dichos grupos, desde el punto de vista de la representación visual, es posible que existan también *aristas sobrantes*, que deberán ser también eliminadas. Para ello, el criterio a seguir será el siguiente: Después de la depuración de una pirámide indexada, cualquier arista que no una un grupo con su predecesor es sobrante y por tanto debe ser también eliminada.

■

REFERENCIAS

[1] **Arcas, A.** y **Cuadras, C.M.** (1987). "Métodos Geométricos de Representación Mediante Modelos en Arbol". *Publicaciones de Bioestadística y Biomatemática.* Universitat de Barcelona.

[2] **Benzecri, J.P.** (1973). *L'Analyse des Données: La Taxonomie. Tome 1.* Dunod. París.

[3] **Bertrand, P.** (1986). "Etude de la Répresentation Pyramidale". *Thèse de Doctorat.* Université de Paris-Dauphine.

[4] **Bertrand, P.** y **Diday, E.** (1985). "A Visual Représentation of the Compatibility Between an Order and a Dissimilarity Index: The Pyramids". *Computational Statistics Quarterly*, **2, Issue 1**, 31–41.

[5] **Capdevila, C.** (1993). "Contribuciones al Estudio del Problema de la Clasificación Mediante Grafos Piramidales". *Tesis de Doctorado en Matemáticas.* Universitat de Barcelona.

[6] **Capdevila, C.** y **Arcas, A.** (1992). "Sobre un Algoritmo de Clasificación Mediante Grafos Piramidales". *Actas de la XX Reunion Nacional de Estadística e Investigación Operativa.* Cáceres 1992.

[7] **Capdevila, C.** y **Arcas, A.** (1993). "Some Aspects About Statistical Inference in Pyramidal Trees". *Comunication n° 41 a la 4th Conference of the International Federation of Classification Societies.* Paris 1993.

[8] **Diday, E.** (1986). "Une Représentation Visuelle des Classes Empietantes: Les Pyramides". *Rairo: Analyse des Données*, **52**, 475–526.

[9] **Diday, E.** (1986). "Ordres and Overlapping Clusters in Pyramides". *Multidimensional Data Analysis.* DSWO Press, Leiden, 201–234.

[10] **Durand, C.** (1986). "Sur la Représentation Pyramidale en Analyse des Données". *Mémoire de DEA en Mathématiques Appliquées.* Université de Provence, Marseille.

[11] **Durand, C.** (1988). "Une Approximation de Robinson Inférieur Maximale". *Rapport de Recherche*, 88–02, Laboratoire de Mathématiques Appliquées et Informatique, Université de Provence, Marseille.

[12] **Durand, C.** y **Fichet, B.** (1988). "One-to-one Correspondences in Pyramidal Representation: A Unified Approach". *Clasification and Related Methods of Data Analysis*. H.H. Bock, Ed. North-Holland. Amsterdam.

[13] **Durand, C.** (1989). "Ordres et Graphes Pseudo-Hierarchiques: Theorie et Optimisation Algoritmique". *Thèse de doctorat en Mathématiques Appliquées*. Universite de Provence. Marseille.

[14] **Fichet, B.** (1984). "Sur une Extension de la Notion de Hiérarchie et son Equivalence Avec Certaines Matrices de Robinson". *Comunication of the Journées de Statistique de Montpellier*.

[15] **Jardine, N.** y **Sibson, R.** (1971). *Mathematical Taxonomy*. Wiley, New York.

[16] **Lerman, I.C.** (1981). *Classification et Analyse Ordinale des Données*. Dunod. Paris.

ENGLISH SUMMARY:

INDEXED PYRAMIDS AND PYRAMIDAL DISSIMILARITIES

Carles Capdevila Marquès and Antoni Arcas Pons

Pyramidal trees, introduced by E. Diday, are a logical generalization of ultrametric trees. They are less restrictive structures where recovering replaces the concept of partition, obtaining a representation which bears more information and is closer to the initial dissimilarities.

The pyramidal representation processes for individuals of a finite population, consists in transforming the initial dissimilarity into a pyramidal dissimilarity by means of some known pyramidal clustering procedure algorithm; this pyramidal dissimilarity is equivalent to an indexed pyramid.

In this study we establish a rigorous mathematic formalization of the equivalence between the concepts of indexed pyramid and pyramidal dissimilarity. In addition, we study the problem of the redundant information contained in a pyramidal structure, and

we provide objective criteria to eliminate it without modificating the corresponding pyramid.

In section 2 you find the basic concepts for the study of the pyramidal models of classification and representation of data, and some elementary relationships among them are established.

We realize that all indexed total hierarchies are also indexed pyramids, and therefore, the pyramidal models constitute a natural generalization of the hierarchical models.

Bearing in mind that the mathematic bases of the hierarchical representations are based on the equivalence between the ultrametric dissimilarities and the indexed hierarchies of parts of a set, a similar equivalence is established between the pyramidal dissimilarities and indexed pyramids of parts of a set.

In section 3.2 it is proved that every pyramidal dissimilarity d defined in Ω, produces an unique indexed pyramid in broad terms of $\wp(\Omega)$, and conversely that all indexed pyramid $(\mathbf{P}, i)$ of $\wp(\Omega)$ induces an unique pyramidal dissimilarity in Ω.

In section 3.3 it is proved that, the indexed pyramid in broad terms of $\wp(\Omega)$ induced by a pyramidal dissimilarity induces the same pyramidal dissimilarity. Conversely, the pyramidal dissimilarity induced by a indexed pyramid in broad terms of $\wp(\Omega)$ induces the same indexed pyramid.

Finally, it is proved that in an indexed pyramid the spare groups can be characterized by: $h \in (\mathbf{P}, i)$ is a *spare group*, if h is a nontrivial intersection of groups of $\mathbf{P}$, and if there exists $h' \in (\mathbf{P}, i)$ such that: $h \subsetneq h'$ and $i(h) = i(h')$.