

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
FACULTAD DE CIENCIAS

LA CLASIFICACION DE LERMAN Y LA
SERIACION MATEMATICA

TESIS
QUE PARA OBTENER EL TITULO DE
MATEMATICO
PRESENTA
MARGARITA ELVIRA CHAVEZ CANO

1979

6693



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

	Pag.
INTRODUCCION	1
CAPITULO I	
SERIACION Y CLASIFICACION	3
1.1 Seriación	3
1.2 Clasificación	8
CAPITULO II	
ALGUNOS RESULTADOS DE I.C. LERMAN EN CLASIFICACION AUTOMATICA	20
2.1 Introducción	20
2.2 Criterios	24
2.2.1 Criterios definidos a partir de la matriz de incidencia de los datos.	24
2.2.2 Criterios definidos a partir de una relación binaria simétrica	27
2.2.3 Criterios definidos a partir del dato de una familia finita de particiones.	35
2.2.4 Criterios definidos a partir del dato de un índice de similitud.	41
2.2.5 Criterios definidos a partir de un preordenamiento	50
CAPITULO III	
ULTRAMETRIA Y SU RELACION CON SERIACION	54
APENDICE	
BIBLIOGRAFIA	

INTRODUCCION

Este trabajo tiene como objetivo, encontrar una relación entre dos ramas de las matemáticas aplicadas, que son: la seriación y la clasificación, esta última tomada desde el punto de vista de Terman. Se pensó que podría existir la relación debido a la forma de las matrices que caracterizan a una seriación y una clasificación respectivamente. Es por esto, que la presentación del material tiene el orden siguiente: el capítulo I comprende, desde el concepto de seriación hasta llegar a una matriz que, bajo ciertas condiciones caracteriza a una seriación; además se da un panorama general de lo que es la clasificación que servirá como introducción al capítulo II, en este último se expone una serie de resultados en clasificación obtenidos por Terman, entre ellos, existe uno en especial, obtenido a través de una cadena de particiones caracterizada por una distancia ultramétrica y poniendo esta distancia en una matriz que será entonces una matriz de distancias ultramétricas. Por último, en el capítulo III se obtiene la relación buscada a través de las matrices mencionadas anteriormente. De esta manera, aunque en general la seriación es utilizada para encontrar el orden secuencial de objetos con respecto al tiempo, se ha llegado a que encontrar una seriación en un conjunto dado de objetos, nos lleva a encontrar una clasificación en el mismo conjunto.

Se presenta también un apéndice en el que se da una serie de resultados en espacios ultramétricos, gracias a los cuales se obtienen los resultados de interés en el capítulo II, este apéndice podría extenderse mucho, lo que constituiría un trabajo completo y el objeto del mismo, es únicamente utilizarlo como herramienta, por lo que se consideran únicamente los resultados que se utilizan en el capítulo mencionado arriba.

CAPITULO I.

SERIACIÓN Y CLASIFICACIÓN.

1.1 SERIACIÓN

La seriación, como muchas ramas de matemáticas aplicadas, no fue inicialmente concebida por matemáticos; ésta se atribuye a Flinders Petrie (1899), pues fue él quien formuló y puso en práctica los principios de lo que él llamó "fechamiento de sucesiones" (sequence-dating), ahora más frecuentemente llamado SERIACION.

La seriación se aplica a un conjunto de objetos para establecer en ellos un orden secuencial, i.e., se busca que entre ellos exista alguna relación de continuidad en algún sentido, por ejemplo el tiempo, teniendo como resultado un orden cronológico. Aquí trataremos solamente el caso en que lo único que se requiere para establecer ese orden, es considerar la presencia o ausencia de algunas características relevantes en los objetos.

La presencia o ausencia de las características se representa por medio de una matriz A , de incidencia, donde los renglones representan los objetos y las columnas las características o variedades. Toda la información queda entonces resumida en esta matriz, cuya definición formal es la siguiente:

Definición 1. Por una matriz de incidencia A se entiende al arreglo:

$$\{a_{ij} : i = 1, \dots, m ; j = 1, \dots, n\},$$

con m renglones y n columnas, donde m es el número de objetos y n es el número de características y

$$a_{ij} = \begin{cases} 1 & \text{si el objeto } i\text{-ésimo posee la característica } j\text{-ésima} \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Entonces la matriz A tiene la forma siguiente:

		CARACTERISTICAS					
		c_1	c_2	c_3	...	c_n	
OBJETOS	o_1	0	1	1	0	...	0
	o_2	1	0	1	1	...	1
	o_3	1	0	0	1	...	0
	·	·	·	·	·	·	·
	·	·	·	·	·	·	·
	·	·	·	·	·	·	·
	o_m	1	0	1	1	...	1

fig. 1

Ahora, si se forma la nueva matriz cuadrada de $n \times n$:

$$c = A^t A,$$

donde, la entrada (i, j) es:

$$c_{ij} = \sum_{k=1}^m a_{ki} a_{kj} .$$

Se puede ver que c_{ij} es igual al número de objetos en los cuales la i -ésima y j -ésima características están presentes.

Por otro lado, si se forma la matriz cuadrada $m \times m$

$$G = AA^t ,$$

cuya entrada (i, j) es:

$$g_{ij} = \sum_{l=1}^n a_{il} a_{jl} ;$$

entonces, g_{ij} es igual al número de características que tienen en común los objetos i y j . Puede decirse entonces que las matrices C y G contienen información acerca del orden secuencial de los datos.

Encontrar una seriación de los objetos, es equivalente a ordenar los renglones de A de manera que en cada columna queden todos los unos juntos, i.e., es equivalente a intercambiar los renglones de forma que el aspecto de la matriz sea como el de la figura 2. De esta forma se tiene cierta continuidad por la manera en que quedan agrupados los unos y queda establecido entonces, el orden secuencial de los objetos.

		CARACTERISTICAS					
		c_1	c_2	c_3	c_4	...	c_n
OBJETOS	o_{10}	1	0	0	0	...	1
	o_8	1	0	1	1	...	1
	o_m	1	1	1	1	...	0
	o_{30}	1	1	1	0	...	0
	...	0	0	0	0	...	0

	o_5	0	0	0	0	...	0
	o_3	0	0	0	0	...	0

Fig. 2

Como puede verse, lo que se obtiene es una nueva matriz de incidencia, a la que Kendall llama matriz de Petrie o matriz P.

La seriación consiste por lo tanto, en poder llevar a una matriz dada de incidencia a una matriz de Petrie, en la cual los renglones rearmados dan el orden secuencial correcto de los objetos.

Fulkerson y Gross (1965), probaron el siguiente resultado:

Si dos matrices de incidencia A_1 y A_2 tienen el mismo número de renglones y el mismo número de columnas, y si $A_1^t A_1 = A_2^t A_2$;

entonces A_2 puede llevarse a la forma de Petrie, i.e., es Petrie-ficible, por permutación de renglones si y sólo si A_1 es Petrie-ficible.

La pregunta "¿es A petrie-ficible?", puede en principio ser contestada si conocemos a la matriz C obtenida de la matriz A ; esto es, puede decirse que la matriz C contiene la información suficiente para decidir si A es petrie-ficible.

Es necesario, en esta parte, decir cuándo una matriz G tiene forma de Robinson: por definición, una matriz G tiene la forma de Robinson si y sólo si, las columnas y los renglones son unimodales y toman su valor máximo sobre la diagonal principal; y se dice que una sucesión de números es unimodal si y sólo si esta sucesión crece (en sentido débil) a un valor máximo y luego decrece (en sentido débil).

David G. Kendall obtuvo el siguiente resultado:

Supóngase que A es petrie-ficible. Entonces la permutación de los renglones que petrie-fica a A es la misma que, cuando se aplica a los renglones y columnas de G simultáneamente, resulta una matriz cuadrada en la forma de Robinson.

Por tanto, después de que se haya decidido que A resulta de la permutación de renglones de

alguna matriz de Petrie; la tarea de llevar a A a la forma de Petrie puede ser realizada utilizando G únicamente.

De las consideraciones anteriores y bajo la hipótesis de que la matriz A de incidencia puede ser llevada a la forma de Petrie, se llega a la siguiente conclusión:

Proposición 1.1

Encontrar una seriación en los objetos es equivalente a ordenar los renglones y columnas de G simultáneamente de manera que resulte una matriz en la forma de Robinson.

1.2 CLASIFICACION

Desde un punto de vista general, una clasificación es la colocación de "individuos" en clases indefinidas inicialmente, de tal forma que los individuos de una clase estén "cercaños" entre sí y "distantes" de los de las otras clases, en algún sentido

Desde el punto de vista de la Biología, taxonomía (del griego $\tau\acute{\alpha}\xi\nu\mu\acute{o}\varsigma$, taxis, orden y $\nu\acute{o}\mu\omicron\varsigma$, nomos, ley) es la ciencia de la clasificación de los organismos. El término taxonomía fue propuesto por el botánico De Candolle para la teoría de

la clasificación de las plantas.

Por lo que hay que hacer notar que en algunas ocasiones los términos "clasificación" y "taxonomía" se usan indistintamente⁽¹⁾. El interés que se persigue en esta sección, es el desarrollo de la teoría matemática de la clasificación, la cual está basada esencialmente en la taxonomía biológica (Jardine & Sibson, 1971; Sneath y Sokal 1973), pudiéndose aplicar a cualquier tipo de datos que satisfagan las propiedades requeridas para el desarrollo de ésta.

La taxonomía tiene tres enfoques, que Sokal & Camin (1965) han distinguido como:

- i) taxonomía operacional,
- ii) taxonomía empírica y
- iii) taxonomía numérica.

De acuerdo a estos autores, la importancia de los enfoques es en el orden dado. En la taxonomía operacional, las proposiciones e hipótesis acerca de la naturaleza pueden ser probadas por observación o experimentación. La taxonomía empírica está basada en la observación de muchas características y de las cuales se puede llevar un récord, aquí se forman grupos de acuerdo a una mayoría de caracteres

(1) En este capítulo así se considerarán.

la clasificación de las plantas.

Por lo que hay que hacer notar que en algunas ocasiones los términos "clasificación" y "taxonomía" se usan indistintamente⁽¹⁾. El interés que se persigue en esta sección, es el desarrollo de la teoría matemática de la clasificación, la cual está basada esencialmente en la taxonomía biológica (Jardine & Sibson, 1971; Sneath y Sokal 1973), pudiéndose aplicar a cualquier tipo de datos que satisfagan las propiedades requeridas para el desarrollo de ésta.

La taxonomía tiene tres enfoques, que Sokal & Camin (1965) han distinguido como:

- i) taxonomía operacional,
- ii) taxonomía empírica y
- iii) taxonomía numérica.

De acuerdo a estos autores, la importancia de los enfoques es en el orden dado. En la taxonomía operacional, las proposiciones e hipótesis acerca de la naturaleza pueden ser probadas por observación o experimentación. La taxonomía empírica está basada en la observación de muchas características y de las cuales se puede llevar un récord, aquí se forman grupos de acuerdo a una mayoría de caracteres

(1) En este capítulo así se considerarán.

compartidos. La taxonomía numérica como generalmente es practicada, es operacional y empírica.

Es importante hacer notar que los taxonomistas numéricos consideran una separación estricta entre la especulación filogenética y el procedimiento taxonómico. Las relaciones taxonómicas se evalúan únicamente en base a los parecidos que existen actualmente entre los objetos o individuos que se tienen a la mano. Estas relaciones fenéticas no toman en cuenta el origen del parecido encontrado ni la razón a la cual los parecidos pueden haberse incrementado o decrementado en el pasado.

Se han hecho estas consideraciones debido a que hasta este momento no se ha hablado de algún esquema de clasificación, y como es de esperarse, cualquier esquema que intente combinar el grado de parecido, descendencia y razón del progreso evolucionario sería excesivamente complicado. Por lo que la separación de las consideraciones fenéticas de las filogenéticas es un avance importante en el pensamiento taxonómico. De acuerdo con Sokal & Sneath, se puede usar únicamente evidencia fenética para establecer una clasificación satisfactoria, las razones por las cuales se asevera esto son las siguientes:

1. El récord del fósil disponible está tan fragmentado que la filogenia de la mayoría de

los taxa es desconocida. Las sucesiones que se ramifican evolucionariamente deben ser probadas extensamente a partir de relaciones fenéticas entre los organismos existentes.

2. La clasificación fenética es posible para todos los grupos. En contraste con la clasificación cladística, basada en sucesiones que se ramifican, que requiere inferencias históricas acerca de la dirección de evolución en un grupo de organismos.

3. Aún cuando la evidencia fósil esté disponible, esta evidencia misma debe ser interpretada primero de una manera fenética estrictamente — con la excepción de que se da además una escala de tiempo, que puede restringir ciertas interpretaciones de la filogenia — ya que los criterios para elegir las formas ancestrales en una filogenia son fenéticas y están basadas en relaciones fenéticas entre supuestos ancestros y descendientes.

4. Desde el punto de vista de la biología en general, probablemente es de más interés describir la similitud total de los organismos que sus sucesiones que se ramifican. Si las clasificaciones son para tener valor predictivo, es evidente que las que están basadas en la similitud total serán las más predictivas.

Entonces, en base a consideraciones técnicas únicamente se dará a continuación la teoría matemática.

Es importante hacer la aclaración que para hacer una clasificación, es esencial suponer que es razonable buscar grupos (o clases) en los datos, siendo éstos subconjuntos del conjunto E de objetos, caracterizados porque poseen propiedades de coherencia y aislamiento.

De aquí en adelante se supondrá dado un conjunto E de "objetos" ⁽¹⁾.

Cuando se desea hacer una clasificación de un conjunto de objetos, la primera idea que se ocurre es partir al conjunto de objetos en subconjuntos, que pueden simplemente formar una cubierta de E (en este caso pueden o no traslaparse los subconjuntos), pueden formar una partición del conjunto E o más aún, se puede tener una sucesión de particiones.

El tipo de clasificación de interés a este trabajo es cuando se tiene una sucesión creciente de particiones; a la que Jardine & Sibson llaman clasificación estratificada jerárquica. Se ha escogido este tipo de clasificación jerárquica ya que posee una estructura más rica que otras clases de agrupamientos.

(1) Se utilizará el término objeto para describir a los elementos del conjunto que está siendo considerado.

La construcción de "sistemas de clasificación"⁽¹⁾ se trata como un proceso que consta de dos partes —la primera parte es construir una medida de disimilaridad (o de similitud) sobre un conjunto de objetos, de tal forma que describa el patrón de variación de los mismos con respecto a los atributos o características escogidos. La segunda parte es representar esta medida (llamada coeficiente de disimilaridad o similitud) por un sistema de clasificación apropiado.

Aquí se supondrá definido un coeficiente de disimilaridad sobre E , el cual es una función real positiva sobre $E \times E$, que satisface las siguientes condiciones:

- i) $d(A, B) \geq 0 \quad \forall A, B \in E$
- ii) $d(A, A) = 0 \quad \forall A \in E$
- iii) $d(A, B) = d(B, A) \quad \forall A, B \in E$.

Sucede frecuentemente que las disimilaridades medidas de alguna manera, tienen ciertas propiedades, por ejemplo, pueden formar una métrica en lugar de un coeficiente de disimilaridad simplemente.

(1) Los términos sistema de clasificación y sistema taxonómico son utilizados para describir cualquier conjunto de subconjuntos de un conjunto de objetos que de alguna manera transmite información acerca de los objetos.

Los valores de un coeficiente de disimilitud para E en general se ponen en un arreglo cuadrado de $n \times n$ (si la cardinalidad de E es igual a n), llamado matriz de disimilaridades, cuyos renglones y columnas están indexados por los objetos de E, como se ilustra en la figura 3.

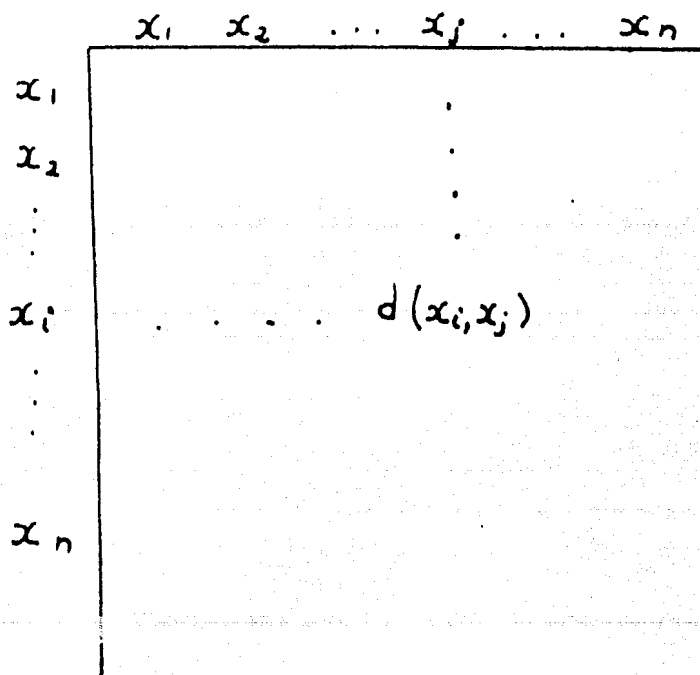


fig. 3

Los métodos jerárquicos de clasificación parten de los coeficientes de disimilaridad (d), i.e., de una matriz de disimilaridades y la transforman en un "dendrograma". El aspecto general de un dendrograma se muestra en la figura 4, en el cual los nodos terminales representan a los objetos (A, B, C, D, E, F, G, H) y la altura de las uniones horizontales (niveles

numéricos) está asociada al parecido entre los objetos de las clases que se unen.

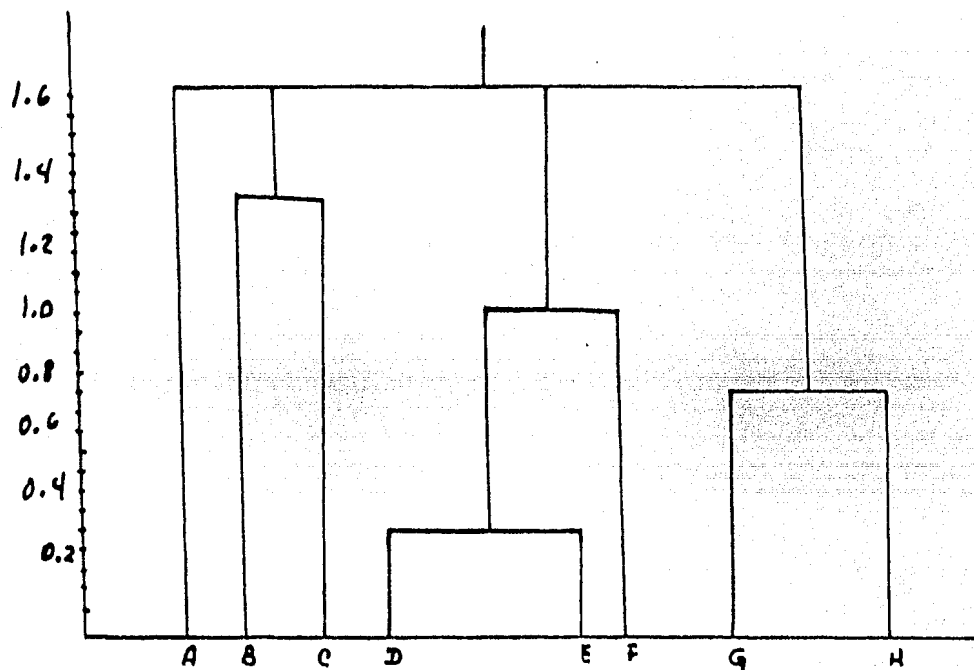


fig. 4

Como puede observarse en la figura 4, los agrupamientos especificados en cada nivel particular de un dendrograma, tienen la propiedad de que son ajenos dos a dos y además, todo elemento de E pertenece a alguna clase (que puede constar de un sólo elemento). De acuerdo con esto, es claro que los agrupamientos forman una partición del conjunto E y como se sabe, existe una correspondencia uno a uno entre las particiones de un conjunto y las relaciones de equivalencia definidas sobre él; donde una relación de equivalencia se define como sigue:

Una relación de equivalencia R sobre E es un subconjunto de $E \times E$, donde se satisfaca que:

- i) $(A, A) \in R \quad \forall A \in E$ (reflexiva)
- ii) $(A, B) \in R \Rightarrow (B, A) \in R \quad \forall A, B \in E$ (simétrica)
- iii) $(A, C) \in R$ y $(C, B) \in R \Rightarrow (A, B) \in R$
 $\forall A, B, C \in E$ (transitiva).

Si R es una relación de equivalencia, entonces los conjuntos:

$$C_A = \{ B : (A, B) \in R \}$$

forman una partición de E , e inversamente, si los conjuntos C_α forman una partición de E , entonces la relación

$$R = \bigcup_{\alpha} C_\alpha \times C_\alpha,$$

es una relación de equivalencia y estas transformaciones de relaciones de equivalencia a particiones y de particiones a relaciones de equivalencia resultan ser, entonces, inversos mutuamente.

Se tiene entonces la siguiente definición formal de un dendrograma:

Sea $\pi(E)$ el conjunto de las relaciones de equivalencia sobre E . Un dendrograma es una función:

$$c: [0, \infty) \longrightarrow \pi(E),$$

que satisface las siguientes condiciones:

- i) $c(h) \leq c(h')$ si $0 \leq h \leq h'$
- ii) $c(h) = E \times E$ para h suficientemente grande
- iii) para h dada, existe $\delta > 0$ tal que:
 $c(h + \delta) = c(h)$.

Si se pide además que todos los elementos de E sean distintos en el nivel cero, a los dendrogramas con esta propiedad se les llama definidos y la propiedad puede ser escrita de la siguiente manera:

$$\text{iv) } c(0) = \Delta E, \text{ donde } \Delta E = \{(A, A), A \in E\}$$

Hasta aquí se ha dado la definición formal de dendrograma y se ha dicho que un método de agrupamiento es una transformación, ya que a los CD's los transforma en dendrogramas. Y aquí surge una pregunta: ¿a partir de los dendrogramas es posible obtener CD's?, de acuerdo con Jardine & Sibson (1971) sí es posible y ellos lo demostraron. A continuación se da un esbozo

de la demostración.

Supóngase que d es un CD arbitrario
y se define:

$$(T_d)(h) = \{ (A, B) : d(A, B) \leq h \}, \quad \forall h \geq 0,$$

es claro que T_d es una relación reflexiva y simétrica pero no es transitiva y por lo que se ha visto, se puede decir que $(T_d)(h)$ es un dendrograma si y sólo si es una relación de equivalencia $\forall h \geq 0$. La condición de transitividad puede ser expresada de la siguiente manera:

$$d(A, c) \leq h \quad \text{y} \quad d(c, B) \leq h \Rightarrow d(A, B) \leq h \quad \forall h \geq 0$$

ó, equivalentemente:

$$\forall A, B, c \in E \quad d(A, B) \leq \max \{ d(A, c), d(c, B) \}$$

Esta condición es conocida como la "Desigualdad Ultramétrica"; por lo que a los CD's que satisfacen esta desigualdad se les llama CD's ultramétricos.

Con el resultado anterior se puede decir que la taxonomía numérica se ha considerado esencialmente, como la colección de métodos que transforman una matriz de disimilaridades en un coeficiente de disimilaridad ultramétrico.

La ultrametricidad es una fuerte restricción pues equivale a que al tomar $A, B, C \in E$, $d(A, B)$, $d(B, C)$ y $d(A, C)$ deben ser las longitudes de los lados de un triángulo isósceles, en el que el lado desigual es, cuando más, tan largo como los otros. En el apéndice A se encuentra una colección de resultados importantes acerca de la ultrametría.

Podemos entonces decir que al transformar una matriz de disimilitudes en un coeficiente ultramétrico se está "deformando" a la matriz original. Es natural pensar que el "mejor" método es aquel que "deforma" lo menos posible a la tabla. Hasta donde se sabe, han habido intentos de precisar qué se entiende por "deformar" y por "mejor" pero no ha habido un acuerdo general, ni un pronunciamiento sólido sobre la conveniencia de usar unas definiciones en vez de las otras en algún tipo de problema.

CAPITULO II

ALGUNOS RESULTADOS DE I. E. LERMAN EN CLASIFICACIÓN AUTOMÁTICA.

2.1 INTRODUCCIÓN.

Todos los resultados que se considerarán en este capítulo suponen dada la información relativa al parecido de los objetos. Esta información generalmente es obtenida de las apariciones sensibles, los caracteres externos de los objetos, sin tomar en cuenta las relaciones posibles entre estos caracteres, i.e., las clasificaciones propuestas son fenotípicas.

Para una clasificación dada, se distinguen en general, dos conceptos de clase:

- i) clase monotética y
- ii) clase politética.

Se dice que una clase es monotética si está caracterizada por una sola propiedad, ésto es, un objeto pertenece a la clase si y sólo si el objeto posee la propiedad. La definición de clase politética, de acuerdo a Beckner (1959), es la siguiente: Una clase K es politética si está asociada a un conjunto J de atributos: a_1, a_2, \dots, a_j tales que:

a) cada elemento de la clase posee una proporción importante (pero no fija) de atributos de J ;

b) cada atributo de J es poseído por una proporción importante de elementos de K ;

c) no existe necesariamente un atributo de J que sea poseído por todos los elementos de K .

Se hace esta consideración ya que al definir únicamente una clase monotética es muy restringido, sobre todo cuando se refiere a las clasificaciones "naturales" dentro de las ciencias naturales y humanas.

Como para la mayoría de los métodos de clasificación, la información de que se dispone está relacionada al parecido de los objetos tomados por parejas y considerando la noción de clase politética, se desea que, con respecto a los atributos de descripción, al grado de parecido entre dos objetos sea grande si pertenecen a la misma clase y pequeño en el caso contrario.

Por lo tanto, se busca sobre el conjunto de objetos, una clasificación (partición) o una jerarquía de clasificaciones (una cadena dentro de la latiz que forman las particiones).

Una cadena de \mathcal{P} , donde \mathcal{P} denota al conjunto de las particiones del conjunto E de objetos, es un subconjunto totalmente ordenado de \mathcal{P} , $\mathcal{C} = (P_1, P_2, \dots, P_k)$ con $P_1 < P_2 < \dots < P_k$ y $P_i \in \mathcal{P}$.

El conjunto \mathcal{P} tiene estructura de látiz ya que, toda pareja de particiones $\{P, P'\}$ tiene un supremo común $P \vee P'$ y un ínfimo común $P \wedge P'$; donde las particiones $P \vee P'$ y $P \wedge P'$ se definen en base a su gráfica como:

$$Gr(P \wedge P') = Gr(P) \cap Gr(P')$$

$$\text{i.e. } x(P \wedge P')y \iff xPy \quad \text{y} \quad xP'y;$$

y $P \vee P'$ se define como: $Gr(P \vee P')$ es la gráfica más pequeña de la relación de equivalencia que contiene al conjunto $Gr(P) \cup Gr(P')$, o equivalentemente: $x(P \vee P')y$ si y sólo si existe una sucesión z_0, z_1, \dots, z_k de elementos de E tales que, para toda $i = 0, 1, \dots, k$

$$z_i P z_{i+1} \quad \text{ó} \quad z_i P' z_{i+1}, \quad \text{con } x = z_0, \quad y = z_k.$$

Una gráfica de una relación de equivalencia P es denotada por $Gr(P)$ y definida como:

$$Gr(P) = \{(x, y) \mid x \in E, y \in E \text{ y } xPy\},$$

donde P también denota la relación de equivalencia.

La naturaleza matemática del problema de la clasificación automática es la siguiente:

Se considera obtenido, de alguna manera, el dato de base (a partir de los objetos), definiendo por ejemplo una relación binaria simétrica, una distancia o pseudo-distancia, etc. Se póngase que a la estructura que forma el dato de base se le llama la estructura de tipo σ sobre el conjunto E de los objetos, esta es menos "rica" que la estructura que se busca, llámese estructura de tipo τ (partición o cadena de particiones). La forma general consiste en relacionar la clase Θ , de las estructuras de tipo τ , sobre E con la clase Σ , de las estructuras de tipo σ . Se trata entonces de determinar el o los elementos de la clase Θ que son tan compatibles como sea posible con el dato de base.

Para establecer lo anterior, se define un CRITERIO; un criterio es, en general, una relación de preorden (que puede ser parcial) sobre el conjunto Θ . Por esta relación se prefiere τ a τ' ($\tau < \tau'$) si, τ es más cercano que τ' a la estructura definida por el dato de base.

El dato de base, a partir del cual se trabaja puede ser:

1. El arreglo $A \times E$ (donde A es el conjunto de los atributos, E de los objetos; $|A| = T$ y $|E| = n$). En este caso Σ es el conjunto de las matrices de incidencia con T filas y n columnas.

2. Una relación binaria simétrica sobre E , entonces Σ es el conjunto de las relaciones binarias simétricas sobre E .

3. Una familia finita de p particiones sobre E , Σ es aquí el conjunto de las familias finitas de p particiones sobre E .

4. Una distancia o pseudo-distancia sobre E (índice de similaridad), Σ es aquí el conjunto de las pseudo-métricas sobre E .

5. Un preorden total sobre E , Σ es entonces el conjunto de los preórdenes totales sobre el conjunto F de las parejas de objetos distintos.

2.2 CRITERIOS

2.2.1 CRITERIOS DEFINIDOS A PARTIR DE LA MATRIZ DE INCIDENCIA DE LOS DATOS.

Ya se definió anteriormente una matriz de incidencia, sin embargo se dará la definición de Lerman: una matriz de incidencia (E_{ij}) $i=1,2,\dots,T$ y $j=1,2,\dots,n$; es el arreglo:

$$E_{ij} = \begin{cases} 1 & \text{si el atributo } a_i \text{ es poseído} \\ & \text{por el objeto } j, \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Aquí, cada objeto está representado por una columna de la matriz (i.e. Lerman considera a la matriz de incidencia transpuesta, en relación a como se definió anteriormente).

Sea (E_1, E_2, \dots, E_k) una partición del conjunto E en k clases; si esta partición define una "clasificación natural", entonces, dada la clase E_i , a ésta le corresponde un subconjunto A_i de atributos, tal que cada caracter de A_i es frecuente en E_i . De acuerdo con esto, los subconjuntos A_i no forman necesariamente una partición, sino en general forman una cubierta, ya que un mismo atributo puede intervenir en la definición de más de una clase. Y lo que se quiere es que esta cubierta sea tal que los traslapes entre las A_i 's sean pequeños; i.e. que:

$$\frac{|A_i \cap A_j|}{|A_i \cup A_j|} \dots \dots \dots 2.2.1$$

Sea pequeña si no es nula. Y entonces, si para un (i, j) dado, $i \neq j$, la razón 2.2.1 es muy grande, no hay razón alguna para que se parta $E_i \cup E_j$ en dos clases.

Definición de un Criterio.

El objetivo, aquí, es encontrar una pareja $(E_i, A_i | i \in I)$, cuyo primer elemento sea una partición sobre E y cuyo segundo elemento sea una familia de subconjuntos de A , que cubran a A , y tal que los elementos de la familia y las clases de la partición se correspondan biyectivamente.

Se denotará por L a la restricción de la matriz de incidencia a $\bigcup_{i \in I} A_i \times E_i$ y se denotará por M a la restricción del complemento.

Se representará a la pareja $(E_i, A_i | i \in I)$ por medio de una matriz π de ceros y unos definida como sigue:

$$\bar{w}_{ij} = \begin{cases} 1 & \text{si el lugar } (i,j) \text{ corresponde a} \\ & \text{un elemento de } L. \\ 0 & \text{si el lugar } (i,j) \text{ corresponde a} \\ & \text{un elemento de } M. \end{cases}$$

$$i = 1, 2, \dots, T, \quad j = 1, 2, \dots, n.$$

Sea $(E_i, A_i | i \in I)$ una pareja formada por una partición sobre E y otra sobre A que se corresponden biyectivamente. Dadas dos parejas, como la anterior, se preferirá aquella para la cual el valor de:

$$\sum_{i \in I} \sum_{j=1}^n (\varepsilon_{ij} - \bar{w}_{ij})^2 \quad \dots \quad 2.2.2$$

sea el más pequeño, donde $\pi = (\bar{w}_{ij})$ es la matriz de ceros y unos relativa a $(E_i, A_i \mid i \in I)$.

La cantidad 2.2.2 representa la restricción de la distancia euclidiana definida en $\mathbb{R}^{T \times n}$ a $\{0,1\}^{T \times n}$, entre el vector que representa a la matriz de incidencia de los datos y la matriz π .

Se puede pedir además, que el número de clases de la partición definida sobre E (que es la partición de interés) sea κ . Con esta restricción, se puede acondicionar el criterio propuesto.

2.2.2. CRITERIOS DEFINIDOS A PARTIR DE UNA RELACION BINARIA SIMETRICA SOBRE E.

En general, la relación se supone no reflexiva. Esta puede ser definida a partir de la elección de un índice de similitud de la manera siguiente:

$$x R y \Leftrightarrow S(x, y) \geq S$$

donde $s(x, y)$ es el índice de similitud ⁽¹⁾
y s es un número fijo dado.

Aquí se representará a una partición por una relación de equivalencia.

Definición de un Criterio.

Se propone un criterio que preordena totalmente al conjunto de las particiones según el valor del cardinal de la diferencia simétrica entre la relación de equivalencia relativa a la partición y R .

Sea (n_1, n_2, \dots, n_k) el tipo ⁽²⁾ de una partición, se puede desarrollar $|R \Delta \pi|$ (el cardinal de la diferencia simétrica) de la forma siguiente:

$$\begin{aligned} |R \Delta \pi| &= |(R \cup \pi) - (R \cap \pi)| \\ &= |R| + |\pi| - 2|R \cap \pi| \quad \dots \quad 2.2.3 \\ &= 2|R - R \cap \pi| - |R| + |\pi| \quad \dots \quad 2.2.4 \end{aligned}$$

Supóngase que el cardinal de la clase E_i es n_i y el de E_j es n_j ; sea entonces r_{ij}

- (1) Las diferentes clases de similitud y disimilitud pueden consultarse en el capítulo 1. de *Mathematical Taxonomy*, Jardine & Sibson 1971
- (2) El tipo de una partición P , es la sucesión decreciente de los cardinales n_j de sus clases.

el cardinal del conjunto de las parejas de la forma (x, y) tales que $x \in E_i, y \in E_j$ y $x R y$;

i.e.,

$$r_{ij} = \left| \{ (x, y) \mid x \in E_i, y \in E_j \text{ y } x R y \} \right|.$$

Teniendo en cuenta que:

$$\sum_{i < j} r_{ij} = 2 \sum_{i < j} r_{ij} \text{ por simetría de la relación, y}$$

$$\sum_i n_i = n; \quad 2.2.4 \text{ puede escribirse como:}$$

$$\begin{aligned} |R \Delta \pi| &= 4 \sum_{i < j} r_{ij} + \sum_i n_i(n_i - 1) - |R| \\ &= 4 \sum_{i < j} r_{ij} + \sum_i n_i(n_i - 1) + 2 \sum_{i < j} n_i n_j - 2 \sum_{i < j} n_i n_j - |R| \\ &= 4 \sum_{i < j} \left(r_{ij} - \frac{n_i n_j}{2} \right) + \sum_i n_i^2 - \sum_i n_i + 2 \sum_{i < j} n_i n_j - |R| \\ &= 4 \sum_{i < j} \left(r_{ij} - \frac{n_i n_j}{2} \right) + \left(\sum_i n_i \right)^2 - n - |R| \\ &= 4 \sum_{i < j} \left(r_{ij} - \frac{n_i n_j}{2} \right) + n^2 - n - |R| \end{aligned}$$

Sea $d_{ij} = r_{ij} - \frac{n_i n_j}{2}$, entonces el crite-

rio preordena al conjunto de las particiones según el valor de:

$$\sum_{i < j} d_{ij}.$$

Como es de esperarse, para este caso existe una partición π que es lo más "cercana" posible a R , la relación que se tiene como dato.

Sea R una relación simétrica fija, sea $m(R)$ el valor mínimo del cardinal de la diferencia simétrica entre R y una relación de equivalencia π ; i.e.

$$m(R) = \min_{\pi \in \mathcal{R}(E)} |R \Delta \pi|,$$

donde $\mathcal{R}(E)$ es el conjunto de las relaciones de equivalencia sobre E .

Nótese que Lerman no considera la reflexividad en la relación considerada, sin embargo el problema de la clasificación hace indispensable esta hipótesis.

En el siguiente lema, Lerman da una buena aproximación, por medio de una cota superior a $m(R)$. La idea intuitiva de cómo se llegó a esa cantidad es la siguiente:

Tratándose siempre de minimizar el cardinal de $R \Delta \pi$ sobre las relaciones de equivalencia π , las particiones extremas (la más fina y la más gruesa) son las que hay que considerar. Sean π_1 y π_2 dichas particiones, entonces:

$$|R \Delta \pi_1| + |R \Delta \pi_2| = \binom{n}{2}.$$

Haciendo variar las particiones, se podría llegar al extremo de que π_1 sea la partición más gruesa y π_2 la más fina, por lo que existe una partición intermedia π de tal forma que las particiones obtenidas antes sean las mismas que se obtienen después de pasar por ésta. Entonces se puede asegurar que se cumple de forma muy aproximada la siguiente igualdad:

$$|R \Delta \pi_1| + |R \Delta \pi_2| = 2 |R \Delta \pi| = \binom{n}{2}$$

De donde:

LEMA.

Se tiene que:

$$m(R) \leq \frac{1}{2} \binom{n}{2}$$

Demostración. Puesto que:

$$|R \Delta \pi_1| + |R \Delta \pi_2| = \binom{n}{2},$$

necesariamente:

$$|R \Delta \pi_1| \leq \frac{1}{2} \binom{n}{2} \quad \text{ó} \quad |R \Delta \pi_2| \leq \frac{1}{2} \binom{n}{2}$$

por lo que:

$$m(R) \leq \frac{1}{2} \binom{n}{2}$$

TEOREMA.

Para todo ε real positivo:

$$\frac{1-\varepsilon}{2} \binom{n}{2} \leq m(R) \leq \frac{1}{2} \binom{n}{2}$$

para casi todas las relaciones R (i.e. para todas las relaciones salvo por una fracción que tiende a cero cuando n tiende a infinito).

Demostración.

Sea F el conjunto de las parejas de objetos diferentes. Sea $f = \binom{n}{2} = |F|$. Sea G_n el cardinal del conjunto de las relaciones de equivalencia que pueden ser definidas sobre E . Entonces:

$$G_n \leq n^n \quad \forall n \in \mathbb{N}.$$

Se comenzará por acotar el cardinal del conjunto de las relaciones R para el cual $m(R) = t$. Para esto, a cada relación de equivalencia π se hace corresponder el conjunto de las relaciones R que están a una distancia t de π , cada una de estas relaciones está definida por la elección de un subconjunto \bar{t} de F de cardinal t en la siguiente forma:

Si p es un elemento cualquiera de \bar{t} entonces $p \in R \Delta \pi$

Si q es un elemento cualquiera de $F - \bar{t}$, entonces $q \in R \cap \pi$.

Tenemos entonces que:

$$|\{R \mid |R \Delta \pi| = t\}| \leq G_m \binom{f}{t}$$

En particular se cumple que:

$$|\{R \mid m(R) = t\}| \leq G_n \binom{f}{t} \leq n^n \binom{f}{t}$$

De donde, una cota del número de relaciones R para las cuales $m(R) < \frac{(1-\epsilon)f}{2}$ está dada por:

$$\sum_{t < \frac{(1-\epsilon)f}{2}} n^n \binom{f}{t} = n^n \sum_{t < \frac{(1-\epsilon)f}{2}} \binom{f}{t} < n^n f \binom{f}{\lfloor \frac{(1-\epsilon)f}{2} \rfloor}$$

Como el cardinal del conjunto de las relaciones R es 2^f , la proporción de aquellas para las cuales $m(R) < \frac{(1-\epsilon)f}{2}$, es estrictamente inferior a:

$$n^n f \binom{f}{\lfloor \frac{(1-\epsilon)f}{2} \rfloor} \frac{1}{2^f}$$

$$\text{pero } 2^f > \binom{f}{\lfloor f/2 \rfloor} \Rightarrow \frac{1}{2^f} < \frac{1}{\binom{f}{\lfloor f/2 \rfloor}}$$

$$\Rightarrow n^n f \binom{f}{\lfloor \frac{(1-\epsilon)f}{2} \rfloor} \frac{1}{2^f} < n^n f \binom{f}{\lfloor \frac{(1-\epsilon)f}{2} \rfloor} / \binom{f}{\lfloor f/2 \rfloor}$$

Ahora:

$$\binom{f}{\lfloor \frac{(1-\epsilon)f}{2} \rfloor} = \frac{f! (f - \lfloor f/2 \rfloor)! \lfloor f/2 \rfloor!}{(f - \lfloor \frac{(1-\epsilon)f}{2} \rfloor)! \lfloor \frac{(1-\epsilon)f}{2} \rfloor! f!}$$

$$\begin{aligned}
&< \frac{[f/2]!}{[(1-\epsilon) \frac{f}{2}]!} \\
&< \sum_{j=0}^{\infty} \binom{-[\epsilon f/2]}{j} (-1)^j \epsilon^j \\
&= \sum_{j=0}^{\infty} \binom{[\epsilon f/2] + j - 1}{j} \epsilon^j \\
&= (1+\epsilon)^{-[\epsilon f/2]}
\end{aligned}$$

Donde: $\lim_{n \rightarrow \infty} (1+\epsilon)^{-[\epsilon f/2]} = 0$

Esto permite asegurar que $n^n f \binom{f}{[(1-\epsilon)f/2]} / 2^f$ tiende a cero cuando n tiende a infinito, y por lo tanto se tiene el resultado deseado.

De acuerdo al teorema anterior para n grande, la distribución de $m(R)$ está muy concentrada inmediatamente a la izquierda de $\frac{1}{2} \binom{n}{2}$, y en consecuencia, si se tiene R una relación simétrica y se obtiene una relación de equivalencia π tal que la distancia entre R y π sea menor o igual a $\frac{1}{2} \binom{n}{2}$ puede considerarse a π como una muy buena aproximación.

2.2.3 CRITERIOS DEFINIDOS A PARTIR DEL DATO DE UNA FAMILIA FINITA DE PARTICIONES.

Los criterios considerados en esta sección se deben a S. Régnier⁽¹⁾ y son estudiados desde el punto de vista de Lerman.

La idea de tener una familia finita de particiones como dato, tiene su origen en considerar que cada uno de los objetos de E puede ser descrito por un cierto número finito de caracteres que pueden corresponder a diversos puntos de vista, la descripción pueda ser definida por una aplicación:

$$E \xrightarrow{a} A_1 \times A_2 \times \dots \times A_h \times \dots \times A_p,$$

donde A_h es el conjunto de los valores del carácter a_h y

$$a = (a_1, a_2, \dots, a_h, \dots, a_p).$$

Se supone que los conjuntos A_h son finitos y que no tienen ninguna métrica natural; de forma que cada carácter a_h define por sí mismo una partición del conjunto E de los objetos, donde cada clase está formada por objetos idénticos desde el punto de vista de a_h ; es decir:

x y y están en la misma clase $\Leftrightarrow a_h(x) = a_h(y)$.

(1) Sur quelques aspects mathématiques des problèmes de la classification automatique. Rapport interne; Centre de Calcul (M. S.H) Paris; et I.C.C. Bull. Vol. 4, p.p. 175-191, 1965.

(En estas condiciones los p caracteres definen p particiones, en general diferentes, del conjunto E).

El dato de una partición P , será representado aquí por el de una familia finita de p particiones idénticas a P .

Definición de un Criterio.

Partición Central. Sobre el conjunto \mathcal{P} de las particiones sobre E se define la distancia d como sigue :

Si $P, P' \in \mathcal{P}$, entonces $d(P, P')$ es el cardinal de la diferencia simétrica de las gráficas en $E \times E$ asociadas respectivamente a las relaciones de equivalencia que definen P y P' .

Esto induce sobre el conjunto S de las familias finitas de p particiones una distancia δ tal que :

$$\delta[(P_1, P_2, \dots, P_p), (P'_1, P'_2, \dots, P'_p)] = \frac{1}{p} \sum_{i=1}^p d(P_i, P'_i) \quad \dots \quad 2.2.5$$

Como se tiene como dato una familia finita de particiones, supóngase que (P_1, P_2, \dots, P_p) es tal familia, entonces el criterio consiste en escoger una partición P tal que la distancia $\delta[(P_1, P_2, \dots, P_p), (P, P, \dots, P)]$ sea la más pequeña; y se llamarán particiones centrales a las que minimicen:

$$\frac{1}{p} \sum_{i=1}^p d(P_i, P) \quad \dots \quad 2.2.6$$

Una forma de simplificar el criterio definido es la siguiente:

Dada la partición P se definen las $(n \times n)$ variables:

$$\bar{w}_{xy} = \begin{cases} 1 & \text{si } x \text{ y } y \text{ están en la misma} \\ & \text{clase} \\ 0 & \text{si no} \end{cases}$$

En consecuencia \bar{w} es la función característica de la gráfica (en $E \times E$) de la relación de equivalencia π asociada a P . \bar{w} es una aplicación de $E \times E$ en el conjunto $\{0, 1\}$, por lo tanto es un elemento de $\{0, 1\}^{E \times E}$.

El conjunto \mathcal{P} de las particiones sobre E es la parte de $\{0, 1\}^{E \times E}$ tal que:

$$P \in \mathcal{P} \Leftrightarrow \begin{cases} \bar{w}_{xy} = \bar{w}_{yx} \\ \bar{w}_{xy} = 1 \text{ y } \bar{w}_{yz} = 1 \Rightarrow \bar{w}_{xz} = 1 \end{cases}$$

Sumergiendo el cubo $\{0, 1\}^{E \times E}$ de todas las relaciones en $\mathbb{R}^{E \times E}$ que es isomorfo a $\mathbb{R}^{n \times n}$, la métrica euclidiana de $\mathbb{R}^{n \times n}$ define una distancia D en el conjunto de las relaciones binarias sobre E , de la forma siguiente:

$$D^2 = (\rho - \rho')^2 = \sum_{(x,y) \in E \times E} (\rho_{xy} - \rho'_{xy})^2 \quad \dots \quad 2.2.7$$

Donde ρ y ρ' son dos relaciones tales que ρ_{xy} y ρ'_{xy} toman valores en $\{0,1\}$. Por lo que \mathcal{D}^2 es el cardinal de la diferencia simétrica de las gráficas de ρ y ρ' en $E \times E$. Si ρ y ρ' definen particiones, entonces:

$$\mathcal{D}^2 = d(\rho, \rho'),$$

i.e., \mathcal{D}^2 es el número de parejas que pertenezcan o satisfacen únicamente a ρ o ρ' , pero no a ambas. Entonces:

$$\frac{1}{P} \sum_{i=1}^P d(P_i, P) = \frac{1}{P} \sum_{h=1}^P \mathcal{D}^2(P_h, P) \quad \dots \quad 2.2.8$$

El objetivo de sumergir al cubo de las relaciones $\{0,1\}^{E \times E}$ en $\mathbb{R}^{E \times E}$ es para poder tener las propiedades de los baricentros, según S. Rágnier.

Se denotará por $S^h = (s_{xy}^h)$ al punto de $\{0,1\}^{E \times E}$ correspondiente a la partición P_h ; $S = \frac{1}{P} \sum_{h=1}^P S^h$ al baricentro de los puntos S^h igualmente ponderados y (\bar{w}_{xy}) el punto asociado a P . Entonces, la expresión 2.2.8 puede escribirse como:

$$\frac{1}{P} \sum_{h=1}^P \left(\sum_{E \times E} (s_{xy}^h - \bar{w}_{xy})^2 \right) = \frac{1}{P} \sum_{h=1}^P \| S^h - \bar{w} \|^2,$$

donde $\|\cdot\|$ es la norma euclidiana definida en $\mathbb{R}^{n \times n}$.

Se puede simplificar aún más la expresión como sigue:

$$\begin{aligned}
 \frac{1}{P} \sum_{h=1}^P \|s^h - \bar{\omega}\|^2 &= \sum_{h=1}^P \frac{1}{P} \|s^h - P\|^2 \\
 &= \sum_{h=1}^P \frac{1}{P} \|s^h - P + S - S\|^2 \\
 &= \sum_{h=1}^P \frac{1}{P} \|(s-P) + (s^h - S)\|^2 \\
 &= \sum_{h=1}^P \frac{1}{P} \|s-P\|^2 + \sum_{h=1}^P \frac{1}{P} \|s^h - S\|^2 \\
 &= \|s-P\|^2 + \frac{1}{P} \sum_{h=1}^P \|s^h - S\|^2
 \end{aligned}$$

Esta descomposición puede ser hecha ya que:

$$\sum_{h=1}^P (s-P) \cdot (s^h - S) = 0$$

Anteriormente se definió a una partición central como aquella para la cual $\frac{1}{P} \sum_{i=1}^P d(P_i, P)$ es mínima, ahora como consecuencia de la simplificación, una partición central puede ser definida como la que realiza el mínimo de $\|s-P\|^2$. Esta función de P se simplifica aún más ya que $\bar{\omega}_{xy} = 0$ ó $1 \neq (x,y)$. Si se construye un punto H tal que todas sus coordenadas sean iguales a $1/2$, entonces el punto P permanezca a una distancia fija del punto H . Entonces:

$$\|P-H\|^2 = \sum_{(x,y) \in E \times E} \left(\omega_{xy} - \frac{1}{2} \right)^2 = \frac{n^2}{4}$$

ya que $|E \times E| = n^2$ y $\left(\omega_{xy} - \frac{1}{2} \right)^2 = \frac{1}{4}$

De donde:

$$\begin{aligned} \|P-S\|^2 &= \|(P-H) - (S-H)\|^2 \\ &= \|P-H\|^2 + \|S-H\|^2 - 2 \langle (P-H), (S-H) \rangle \\ &= \frac{n^2}{4} + \|S-H\|^2 - 2 \langle (P-H), (S-H) \rangle. \end{aligned}$$

Por lo tanto, encontrar el minimo de $\|P-S\|$ es equivalente a maximizar el producto escalar:

$$\langle P-H, S-H \rangle \quad (\text{dado } S)$$

El cual puede escribirse de la siguiente forma:

$$\langle P-H, S-H \rangle = \sum_{(x,y) \in E \times E} \left(\bar{\omega}_{xy} - \frac{1}{2} \right) \left(S_{xy} - \frac{1}{2} \right),$$

donde S_{xy} son las coordenadas del punto S , i.e.,

$$S_{xy} = \frac{1}{p} \sum_{h=1}^p S_{xy}^h.$$

Sea $t_{xy} = S_{xy} - \frac{1}{2}$, así una partición central es la que maximiza la forma lineal:

$$L(p) = \sum_{(x,y) \in E \times E} t_{xy} \bar{\omega}_{xy} \quad \dots \quad 2.2.9$$

Como una partición P sobre E define las clases E_1, E_2, \dots, E_k ; entonces 2.2.9 representa:

$$\sum_{h=1}^k \left(\sum_{E_h \times E_h} t_{xy} \right)$$

que se puede considerar como una medida de la gráfica de la relación de equivalencia definido por P . Si se considera a t_{xy} como la masa del vértice (x, y) de la gráfica, entonces

$$\sum_{E_h \times E_h} t_{xy},$$

es una forma de compacidad de la clase E_h y $L(p)$ es la suma de las compacidades de las diferentes clases.

2.2.4 CRITERIOS DEFINIDOS A PARTIR DEL DATO DE UN INDICE DE SIMILARIDAD.

Se supone definido un índice de similitud en E , que es una función real positiva sobre $E \times E$, $S(x, y)$, considerada como medidor de los parecidos entre objetos. S es, entonces, un elemento de $\mathbb{R}_+^{E \times E}$, que es isomorfo a $\mathbb{R}^{n \times n}$.

Los valores de un índice de similitud para E , en general se ponen en un arreglo cuadrado de $n \times n$, cuyos renglones y columnas son indexados por los objetos de E , como se muestra en la figura 5

Definición de un criterio.

Sin perder generalidad, casi siempre se puede restringir al caso donde $S(x, y)$ toma valores en el intervalo $[0, 1]$ y donde $S(x, x) = 1 \quad \forall x \in E$.

	x_1	x_2	\dots	x_i	\dots	x_n
x_1						
x_2						
\vdots						
x_j				$s(x_j, x_i)$		
\vdots						
x_n						

fig. 5

Sea S un elemento de $\mathbb{R}^{n \times n}$ que represente al índice de similaridad y P el elemento de $\{0,1\}^{E \times E}$ asociado a una partición mediante la introducción del cubo de relaciones en $\mathbb{R}^{n \times n}$; el objetivo que se persigue es encontrar una partición P tal que $\|P - S\|$ sea muy pequeño y donde $\|\cdot\|$ es la norma euclidiana en $\mathbb{R}^{n \times n}$.

Ahora, considerando que el número de clases puede ser fijado a priori y que se puede trabajar en un espacio con un producto escalar, se presenta un criterio creado por Vo Khac Khoan y Ngliem Phong Tuan.

Se asocia a cada clase E_i su vector medio:

$$E_i \longrightarrow \vec{g}(E_i) = \frac{1}{|E_i|} \sum_{\vec{x} \in E_i} \vec{x}.$$

Se define la dispersión interna en E_i por:

$$\Delta(E_i) = \sum_{\vec{x} \in E_i} \|\vec{x} - \vec{g}(E_i)\|^2.$$

Con respecto a las diferentes clases, se define la dispersión compuesta como sigue:

$$\Delta(E_1, E_2, \dots, E_k) = \sum_{i=1}^k \|\vec{g}(E_i) - \vec{G}\|,$$

donde \vec{G} es el vector medio de la población.

El criterio consiste en considerar una partición de tal forma que la dispersión compuesta de las clases involucradas en ésta, sea grande. Pero maximizar la dispersión compuesta de (E_1, E_2, \dots, E_k) se reduce a minimizar la suma de las dispersiones internas de las diferentes clases.

Criterio para encontrar una Cadena de Particiones.

Siendo el dato una métrica o pseudo-métrica:

$$D'(x, y) = 1 - S(x, y),$$

(donde $S(x, y)$ es un índice de similaridad con valores en $[0, 1]$), se representará a una cadena de particiones por el dato, sobre E , de una distancia ultramétrica de acuerdo a la proposición 5 del apéndice.

Refiriéndose al dato de una matriz de distancias, M. Eytan caracteriza a las que corres-

ponden a una distancia ultramétrica en la forma en que aparece en la proposición 2, para cuya demostración se da el siguiente resultado:

Proposición 1. Dado un coeficiente ultramétrico de disimilitud definido en E , puede construirse un orden lineal en E tal que las clases definidas por los α -cortes, son intervalos.

Demostración. por inducción sobre $n = |E|$.

i) Para $n = 2$, es obvio.

ii) Paso inductivo. Sean $x, y \in E$ tales que $d(x, y) \leq d(z, w) \forall z, w \in E$.

Definimos $E' = (E - \{x, y\}) \cup \{xy\}$ y

$$d'(z, w) = \begin{cases} d(z, w) & \text{si } z \neq x, w \neq y \\ d(x, w) & \text{si } z = xy \text{ y } w \neq y \\ d(z, x) & \text{si } w = xy, z \neq x \\ 0 & \text{si } z = w = xy \end{cases}$$

Lema. d' es un coeficiente de disimilitud ultramétrico.

Demostración. Basta demostrar que todo triángulo es isósceles y que el lado desigual es menor o igual que los otros.

Sean $z, w, t \in E$. Si ninguno es xy , ultrametría de d implica lo deseado.

ponden a una distancia ultramétrica en la forma en que aparece en la proposición 2, para cuya demostración se da el siguiente resultado:

Proposición 1. Dado un coeficiente ultramétrico de disimilaridad definido en E , puede construirse un orden lineal en E tal que las clases definidas por los d -cortes, son intervalos.

Demostración. por inducción sobre $n = |E|$.

i) Para $n=2$, es obvio.

ii) Paso inductivo. Sean $x, y \in E$ tales que $d(x, y) \leq d(z, w) \forall z, w \in E$.

Definimos $E' = (E - \{x, y\}) \cup \{xy\}$ y

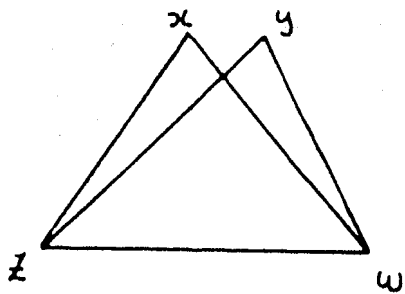
$$d'(z, w) = \begin{cases} d(z, w) & \text{si } z \neq x, w \neq y \\ d(x, w) & \text{si } z = xy \text{ y } w \neq y \\ d(z, x) & \text{si } w = xy, z \neq x \\ 0 & \text{si } z = w = xy \end{cases}$$

Lema. d' es un coeficiente de disimilaridad ultramétrico.

Demostración. Basta demostrar que todo triángulo es isósceles y que el lado desigual es menor o igual que los otros.

Sean $z, w, t \in E$. Si ninguno es xy , la ultrametría de d implica lo deseado.

Sea $t = xy$



Si el triángulo no cumple lo deseado, puede construirse otro (con x ó y como tercer vértice) que tampoco lo cumpliría para d , lo cual es imposible.

Ahora, por hipótesis de inducción (ya que $|E'| = |E| - 1$), para E' existe el orden con la propiedad enunciada y por lo tanto también para E .

En este punto hay que aclarar que Jardine y Sibson (Jardine & Sibson 1971), han probado que existe una correspondencia biunívoca entre el conjunto de dendrogramas y el de coeficientes ultramétricos de disimilitud (distancias ultramétricas).

Proposición 2. Para que una matriz (de $n \times n$) de números reales positivos, simétrica con respecto a la diagonal, sea una matriz de distancias asociada a una distancia ultramétrica sobre E , es necesario y suficiente que

se pueda llevarla por permutación de renglones y columnas del mismo número a la forma siguiente:

a) alejándose del término diagonal, que vale cero, los elementos de un mismo renglón van creciendo en valor;

b) para todo k , si $d(k, k+i) \neq d(k, k+i+1)$ entonces $d(k, k+j) = d(k+i, k+j)$ $j = i+2, i+3, \dots$

Demostración.

b) i) $d(k, k+i) < d(k, k+i+1)$ $i = 1, 2, 3, \dots$

ya que por la proposición 1, si $d = d(k, k+i+1)$ entonces el intervalo $[k, k+i+1]$ está en una clase, por lo tanto, $d(k, k+i) \leq d = d(k, k+i+1)$, pero como por hipótesis $d(k, k+i) \neq d(k, k+i+1)$ se tiene entonces el resultado deseado.

ii) $d(k, k+i) < d(k, k+j)$, $j = i+2, i+3, \dots$

esta afirmación puede demostrarse de la siguiente manera: de i) se tiene que

$$d(k, k+i) < d(k, k+i+1)$$

Sea $d = d(k, k+j)$ $j = i+2, i+3, \dots$ y supongamos que $d(k, k+i) = d$. Entonces, por la proposición 1, el intervalo $[k, k+i+1]$ está contenido en una clase a nivel d y por lo tanto:

$$d(k, k+i+1) \leq d = d(k, k+i) !!!$$

Ahora, consideremos $d(k, k+j)$ y $d(k+i, k+j)$ para $j = i+2, i+3, \dots$

Por (c) se sabe que

$$d(k, k+j) > d(k, k+i) \quad j=i+2, i+3, \dots$$

por lo tanto:

$$d(k, k+j) = d(k+i, k+j) \quad j=i+2, i+3, \dots$$

debido a que se cumple la ultrametría.

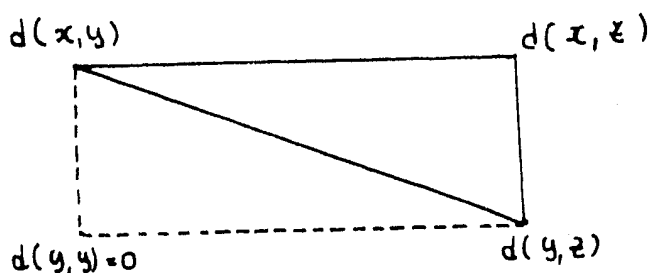
a) para el caso en que $d(k, k+i) \neq d(k, k+i+1)$, queda demostrado a) de acuerdo al inciso c) de b). El otro caso incluye la posibilidad de que $d(k, k+i) = d(k, k+i+1)$ lo cual, de manera análoga a la demostración del inciso c) de b) se puede concluir que $d(k, k+i) \leq d(k, k+i+1)$.

Inversamente, sea d una matriz simétrica en la cual se verifican (a) y (b). La condición de simetría asegura todos los axiomas excepto el de ultrametría (UM). Entonces, basta verificar que (a) y (b) implican (UM). Para esto, consideremos a los elementos x y z de E fijos, entonces hay que verificar que para todo $y \in E$, (UM) es verdadero.

Se considera un orden entre los elementos de E definido por d . Entonces puede suceder que un elemento y sea interior o exterior al intervalo:

$$(x, z) = \{ y : x < y < z \}$$

1^{er} caso. y es interior al intervalo. Consideremos al triángulo:



(i) Si $d(x, y) \neq d(x, z)$, esto es, por el inciso a), si $d(x, y) < d(x, z)$, entonces por b) se tiene que $d(y, z) = d(x, z)$. Por lo tanto:

$$d(x, z) = \max \{ d(x, y), d(y, z) \} = d(y, z)$$

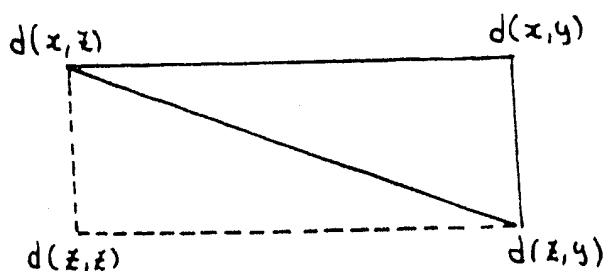
(ii) Si $d(x, y) = d(x, z)$, entonces por

$$(b) \quad d(y, z) \leq d(x, z)$$

Por lo tanto:

$$d(x, z) \leq \max \{ d(x, y), d(y, z) \} = d(x, y)$$

2^o caso. y es exterior al intervalo ($x < z < y$). Consideremos al triángulo:



Por a) se tiene que:

$$d(x, z) \leq d(x, y)$$

Por lo tanto:

$$d(x, z) \leq \max \{ d(x, y), d(y, z) \} = d(x, y).$$

Definición de un Criterio.

Se desea encontrar las matrices ultramétricas Δ que minimicen la distancia, en el sentido de la métrica euclidiana en $\mathbb{R}^{n \times n}$ entre Δ y la matriz dada D de distancias (o pseudo-distancias) sobre E , i.e. se busca

$$\| \Delta - D \|^2 \text{ mínimo.}$$

Ya que las matrices $D = (d_{ij})$ y $\Delta = (\delta_{ij})$ son simétricas y además $d_{ii} = \delta_{ii} = 0 \forall i$, el problema se reduce a minimizar:

$$\sum_{i < j} (\delta_{ij} - d_{ij})^2$$

bajo las restricciones:

$$\delta_{ij} \leq \max \{ \delta_{ik}, \delta_{kj} \} \quad \forall i, j, k$$

El problema así planteado es muy complejo, y de la misma manera es difícil concebir un algoritmo que opere por permutación de las hileras y columnas de D respectivamente asociadas, llegando a un óptimo local coherente con el criterio conocido.

C. J. Jardine, N. Jardine y R. Sibson definen sobre el conjunto de las métricas sobre E la relación de "dominancia"

(orden) siguiente:

Si d y d' son coeficientes de disimilitud se dice que d domina a d' ($d \succ d'$) si y sólo si $d(x, y) \geq d'(x, y) \quad \forall x, y \in E$ y se prefiere la distancia ultramétrica δ que realiza el $\sup \{ \delta \mid \delta \in \Delta \text{ y } \delta \leq d \}$, donde Δ denota al conjunto de las distancias ultramétricas y d la métrica definida sobre E . Este procedimiento ya está bien estudiado y corresponde al método de conexión simple (single link), que es una función continua de los datos (ver Jardine & Sibson 1973).

2.2.5 CRITERIOS DEFINIDOS A PARTIR DE UN PREORDENAMIENTO SOBRE E .

Con frecuencia se presenta el problema de construir una clasificación (en el sentido de partición) para una familia E de objetos, a partir de un coeficiente de similitud S definido para las parejas de ellos.

En esta sección se presentan dos criterios cuya optimización conduciría a la construcción de esa clasificación. Con este propósito se introducen las siguientes definiciones:

Sea F el conjunto de las parejas de objetos de E . Sea ρ una relación de equivalencia en E , y sean

$$R = \{(x, y) \mid x \rho y\} \quad \text{y} \quad T = F - R.$$

Sea W la gráfica correspondiente al preorden total en $F \times F$ definido como sigue:

$$(x, y) \in (z, t) \Leftrightarrow S(x, y) \gg S(z, t).$$

Criterio de J. P. Benzecri:

J. P. Benzecri propone preferir a una partición si el cardinal de la intersección en $F \times F$ de W y $R \times S$ es el más grande. De tal forma que el problema de la búsqueda de una partición (clasificación) se reduce a maximizar $|W \cap R \times S|$, donde R y S son relativos a una relación de equivalencia ρ .

Criterio de W. F. de la Vega

Para la búsqueda de una clasificación, W. F. de la Vega propone maximizar el cardinal del complemento en $F \times F$ de la diferencia simétrica entre W y $R \times S$, esto es, maximizar $|F \times F - W \Delta R \times S|$.

Como:

$$|W \cup R \times S| = |W| + |R \times S| - |W \cap R \times S|,$$

y por otro lado:

$$\begin{aligned}
 |W \Delta R \times S| + |W \cap R \times S| &= \\
 &= |(W \cup R \times S) - (W \cap R \times S)| + |W \cap R \times S| \\
 &= |W \cup R \times S|
 \end{aligned}$$

De donde se obtiene:

$$\begin{aligned}
 |F \times F - W \Delta R \times S| &= |F \times F| - |W \Delta R \times S| \\
 &= |F|^2 - |W \cup R \times S| + |W \cap R \times S| \\
 &= |F|^2 - (|W| + |R \times S| - |W \cap R \times S|) + |W \cap R \times S| \\
 &= |F|^2 - |W| - |R \times S| + 2|W \cap R \times S| \\
 &= |F|^2 - |W| + 2(|W \cap R \times S| - \frac{1}{2}|R| \times |S|)
 \end{aligned}$$

Y ya que el objetivo es maximizar $|F \times F - W \Delta R \times S|$, el problema se reduce a maximizar:

$$|W \cap R \times S| - \frac{1}{2}|R| \times |S|,$$

donde W está dado y donde R y S son relativos a una misma partición P (buscada).

Aquí también se puede trabajar con formas lineales definidas sobre el espacio $\{0,1\}^{F \times F}$ si se introducen las funciones indicadoras I_W y $I_{R \times S}$ de los subconjuntos W y $R \times S$ de $F \times F$, entonces el criterio de J. P. Benzécri se puede expresar de la siguiente forma:

$$\max \sum_{F \times F} I_{R \times S}(p, q) I_W(p, q)$$

y al segundo :

$$\max \sum_{F \times F} I_{R \times S}(p, q) \left[I_W(p, q) - \frac{1}{2} \right]$$

Las dos expresiones anteriores son formas lineales definidas sobre el espacio $\{0, 1\}^{F \times F}$.

CAPITULO III

ULTRAMETRIA Y SU RELACION CON SERIACION.

En el capítulo II se han expuesto algunos resultados de Lerman en clasificación, entre estos, el obtenido en la sección 2.4 es el de principal interés aquí.

En la proposición 2 de la sección 2.4, capítulo II, se caracteriza a una matriz para que ésta sea una matriz de distancias ultramétricas y en el criterio definido para encontrar una clasificación (en el sentido de partición) se trata de obtener una matriz de distancias ultramétricas Δ tal que:

$$\| \Delta - D \|^2$$

sea mínimo y donde D es el dato de base (matriz construida a partir de un índice de similitud). Por la proposición 5 del apéndice, se sabe que a la matriz de distancias ultramétricas está asociada canónicamente una cadena \mathcal{P} , i.e., una sucesión creciente de particiones de E ; por lo que al encontrar una matriz Δ con la característica anterior, se ha encontrado una clasificación en E .

Consideremos ahora la proposición 1.1 del capítulo I, donde queda caracterizada una

seriación a través de una matriz de Robinson, la cual se definió anteriormente y que por la forma en que queda construida puede pensarse en ella como una matriz de similitudes.

Se pensó que habría cierta relación entre una matriz de distancias ultramétricas y una matriz de Robinson debido a la forma "dual" en que están caracterizadas, de las cuales, tener una matriz como la primera es equivalente a haber encontrado una clasificación de los datos, y tener una matriz como la segunda es equivalente, bajo ciertas condiciones, a tener una seriación de los datos. La relación es entonces entre clasificación y seriación, a través de dichas matrices y que se presenta en el siguiente teorema.

TEOREMA.

Sea d_{ij} el elemento (i, j) de una matriz D (de distancias ultramétricas), entonces

$$r_{ij} = \frac{d_{i1} + d_{j1} - d_{ij}}{2}$$

es el elemento (i, j) de una matriz en forma de Robinson.

Demostración. Para demostrar que la matriz R obtenida es de Robinson, basta probar que para $i \leq j$ se tiene que:

$$r_{ij} \gg r_{i,j+1} \quad \text{y} \quad r_{nn} \gg r_{nj}$$

Entonces hay que verificar primero que:

$$r_{ij} = \frac{d_{i1} + d_{j1} - d_{ij}}{2} \gg \frac{d_{i1} + d_{j+1,1} - d_{i,j+1}}{2} = r_{i,j+1}, \quad i \leq j$$

lo cual es equivalente a que se cumpla:

$$d_{j1} - d_{ij} \gg d_{j+1,1} - d_{i,j+1}, \quad i \leq j \dots 3.1$$

Se considerará primero el caso en que $i=j$.

Como d es una métrica, en particular cumple con la desigualdad del triángulo:

$$d_{j+1,1} \leq d_{j+1,j} + d_{j,1}$$

$$\Leftrightarrow d_{j+1,1} \leq d_{j,j+1} + d_{j,1} \quad \text{ya que } d_{j+1,j} = d_{j,j+1}$$

$$\Leftrightarrow d_{j+1,1} - d_{j,1} \leq d_{j,j+1}$$

$$\Leftrightarrow d_{j+1,1} - d_{j,1} \leq d_{j,j+1} - d_{jj} \quad \text{ya que } d_{jj} = 0$$

$$\Leftrightarrow d_{j,1} - d_{j+1,1} \gg d_{jj} - d_{j,j+1}$$

$$\Leftrightarrow d_{j,1} - d_{jj} \gg d_{j+1,1} - d_{j,j+1}$$

lo cual prueba 3.1 cuando $i=j$.

Para el caso $i < j$ se demostrará que :

$$d_{i,j+1} - d_{ij} \leq d_{i,j+1} - d_{i,j}$$

que es equivalente a 3.1, lo cual quedará probado si se verifica que:

$$d_{i-1,j+1} - d_{i-1,j} \leq d_{i,j+1} - d_{i,j} \quad i < j$$

Por la hipótesis b) de la proposición 2, sección 2.4, capítulo II :

Si $d_{i-1,i} \neq d_{i-1,i+1}$ entonces:

$$d_{i-1,j} = d_{i,j} \quad \forall i < j$$

$$\therefore d_{i-1,j+1} - d_{i-1,j} = d_{i,j+1} - d_{i,j} \quad \forall i < j$$

$$\therefore d_{i-1,j+1} - d_{i-1,j} \leq d_{i,j+1} - d_{i,j} \quad \forall i < j$$

Ahora, si $d_{i-1,i} = d_{i-1,i+1} = \dots = d_{i-1,i+l}$

entonces:

$$i) \quad d_{ij} \leq d_{i-1,j} \quad \text{para } i < j \leq i+l-1$$

$$ii) \quad d_{ij} = d_{i-1,j} \quad \text{para } j > i+l$$

$$ii) \quad d_{ij} = d_{i-1,j} \Rightarrow d_{ij} \leq d_{i-1,j} \quad \dots \quad 3.2$$

$$d_{ij} = d_{i-1,j} \Rightarrow d_{i,j+1} = d_{i-1,j+1}$$

$$\Rightarrow d_{i-1,j+1} \leq d_{i,j+1} \quad \dots \quad 3.3$$

De 3.2 y 3.3 se tiene que

$$d_{i-1,j+1} - d_{i-1,j} \leq d_{i,j+1} - d_{i,j}$$

i) Se sabe que:

$$d_{ij} \leq d_{i-1,j} \quad \text{para } i < j < i+1 \dots 3.4$$

Ahora, como d es una ultramétrica se tiene que

$$d_{j+1,i-1} \leq \max \{ d_{j+1,i}, d_{i,c-1} \},$$

donde:

$$d_{i,c-1} \leq d_{i,j+1} = d_{j+1,i}$$

ya que $j > i \Rightarrow j+1 > i-1$ y por la hipótesis

a), proposición 2, sección 2.4, capítulo II

$$\therefore \max \{ d_{j+1,i}, d_{i,c-1} \} = d_{j+1,i} \dots 3.5$$

\therefore de 3.4 y 3.5 se obtiene:

$$d_{i-1,j+1} - d_{i-1,j} \leq d_{i,j+1} - d_{i,j}.$$

APENDICE.

Espacios Ultramétricos.

Propiedades:

Sea (E, d) un espacio métrico, i.e. un conjunto E y una función real positiva d

$$d: E \times E \longrightarrow \mathbb{R}^+,$$

tal que:

$$(M1) \quad d(x, y) = d(y, x) \quad \forall x, y \in E$$

$$(M2) \quad d(x, y) = 0 \quad \text{si y solo si} \quad x = y \\ \forall x, y \in E$$

$$(M3) \quad d(x, z) \leq d(x, y) + d(y, z).$$

Se dice que (E, d) es un espacio ultramétrico si d verifica además que:

$$(UM) \quad d(x, z) \leq \max(d(x, y), d(y, z)) \\ \forall x, y, z \in E.$$

Generalmente se le llama a esta desigualdad, la "desigualdad ultramétrica".

Definición. Un divisor de E es una relación de equivalencia D en E tal que:

$$\forall a, b, x, y \in E \\ a D b \ \& \ (d(x, y) \leq d(a, b)) \Rightarrow x D y.$$

Teorema 1. Un espacio métrico (E, d) es ultramétrico si y sólo si todos los triángulos son isósceles y la longitud de la base es menor o igual a la longitud de los lados.

Demostración.

Supongamos primero que (E, d) es un espacio ultramétrico y sea x, y, z un triángulo en él, donde se supone que x, y es el lado más chico. Entonces:

$$d(x, z) \leq \max \{ d(x, y), d(y, z) \} = d(y, z),$$

permutando x y y se obtiene:

$$d(y, z) \leq \max \{ d(y, x), d(x, z) \} = d(x, z)$$

$$\Rightarrow d(x, z) = d(y, z) \geq d(x, y)$$

$\therefore x, y, z$ es un triángulo isósceles.

Ahora, suponiendo que todos los triángulos son isósceles, esto implica directamente la propiedad ultramétrica.

Proposición 1. Todo punto de un círculo en un espacio ultramétrico E , es centro.

Demostración.

Considérese un círculo con centro a y radio r y sea b un punto en ese círculo, entonces se tiene que demostrar que:

$$C(a, r) = \{x \mid d(a, x) \leq r\} = \{x \mid d(b, x) \leq r\} = C(b, r)$$

Por hipótesis $b \in C(a, r) \Rightarrow d(a, b) \leq r$

Ahora sea $x \in C(a, r) \Rightarrow$

$$d(b, x) \leq \max \{d(a, b), d(a, x)\} \leq r$$

$$\Rightarrow C(a, r) \subset C(b, r)$$

Sea $x \in C(b, r) \Rightarrow$

$$d(a, x) \leq \max \{d(a, b), d(b, x)\} \leq r$$

$$\Rightarrow C(b, r) \subset C(a, r)$$

$$\therefore C(a, r) = C(b, r).$$

Corolario 1.1. Dos círculos no ajenos, en un espacio ultramétrico, son concéntricos

Demostración. Sean $C(a, r)$ y $C(b, r')$ tales que $C(a, r) \cap C(b, r') \neq \emptyset$ y $r \leq r'$. Sea $x \in C(a, r) \cap C(b, r')$, entonces por la proposición 1 x es centro de $C(a, r)$ y es centro de $C(b, r')$. Por lo tanto $C(a, r)$ y $C(b, r')$ son concéntricos.

Corolario 1.2. Dos círculos del mismo radio no ajenos, en un espacio ultramétrico, coinciden.

Demostración. Consecuencia inmediata del corolario 1.

Corolario 1.3. Los círculos del mismo radio, forman una partición del espacio ultramétrico E , la equivalencia correspondiente es un divisor de E .

Demostración.

Consecuencia directa de las definiciones de partición y de divisor.

Proposición 2. Sean e y e' dos círculos ajenos en E , sean x y x' en e y y y y' en e' , entonces:

$$d(x, y) = d(x', y').$$

Demostración. Considere los triángulos x, y, y y x', y, y' los cuales son isósceles de acuerdo al teorema 1. Se tiene entonces que:

$$d(x, y) = d(x, y') = d(x', y) = d(x', y').$$

Proposición 3. Todo círculo de radio r que no contiene a su frontera, en E y cuya intersección con un círculo de radio r en E es no vacía, está contenido en éste.

Demostración.

Sea $e(a, r^-) = \{x \mid d(a, x) < r\}$ el círculo sin frontera y sea $e(b, r) = \{x \mid d(b, x) \leq r\}$ tal que $e(a, r^-) \cap e(b, r) \neq \emptyset$

Sea $x \in e(a, r^-) \Rightarrow d(a, x) < r$, entonces hay que demostrar que $d(b, x) \leq r$.

Sea $y \in e(a, r^-) \cap e(b, r) \Rightarrow$

$$d(a, y) < r \quad \text{y} \quad d(b, y) \leq r$$

Considerando el triángulo a, b, y se tiene:

$$d(a, b) \leq \max \{d(a, y), d(y, b)\} \leq r$$

Entonces:

$$d(b, x) \leq \max \{d(b, a), d(a, x)\} \leq r$$

$$\therefore d(b, x) \leq r \quad \therefore x \in e(b, r)$$

$$\therefore e(a, r^-) \subset e(b, r)$$

Proposición 4. Sea D un divisor de E , con E finito, sea $\delta = \max \{d(a, b) \mid a \triangleright b\}$ y sea Δ una relación de equivalencia en E tal que $a \Delta b \Leftrightarrow d(a, b) \leq \delta$. Entonces, todos los divisores de E son de la misma forma que Δ para alguna δ .

Demostración.

Sean $\bar{a}, \bar{b}, x, y \in E$.t. $d(\bar{a}, \bar{b}) = \delta$,

$$\bar{a} \mathcal{D} \bar{b} \text{ y } d(x, y) \leq \delta = d(\bar{a}, \bar{b}) \Rightarrow x \mathcal{D} y$$

$$\therefore \text{ si } x \Delta y \Rightarrow x \mathcal{D} y \text{ y } \Delta \in \mathcal{D}.$$

$$\text{Ahora, sean } x, y \in E \text{ tales que } x \mathcal{D} y \Rightarrow d(x, y) \in \delta \Rightarrow x \Delta y \text{ y } \mathcal{D} \in \Delta$$

$$\therefore \mathcal{D} = \Delta.$$

Corolario 4.1. Supóngase que E es finito, entonces el número de δ 's diferentes tales que generan divisores distintos es $|Im(E \times E)|$ (el cardinal de la imagen de $E \times E$).

Proposición 5. A todo espacio ultramétrico finito (E, d) , está asociada una sucesión $\{P_i\}_{i \in I}$ finita, creciente de particiones de E , tal que los P_i son todos los divisores de E .

Demostración.

Por el corolario 3 de la proposición 1, es suficiente tomar los círculos de radio i y de hacer recorrer i en \mathbb{N} y ya que E es finito, i recorrerá solamente un segmento finito I de \mathbb{N} .

Proposición 6. Sea $\{P_i\}_{i \in I}$ una sucesión finita creciente de particiones de E . Sea $d: E \times E \rightarrow I \subset \mathbb{N}$ una función definida como:

$$d(x, y) = \min \{ i \mid (x, y) \in P_i \}$$

Entonces:

- i) d es una distancia ultramétrica y
- ii) $\{P_i\}_{i \in I}$ es el conjunto de divisores de E .

Demostración.

i) d es reflexiva y simétrica obviamente, $d(x, z), d(z, y) \in P_{\max \{d(x, z), d(z, y)\}}$

porque $\{P_i\}$ es una sucesión creciente, como $P_i \in \mathcal{P}, i \in I$, $(x, y) \in P_{\max \{d(x, z), d(z, y)\}}$
 $\therefore d(x, y) \leq \max \{d(x, z), d(z, y)\}$

ii) Sea $i \in I$ y sea $\Delta_i: E \times E \rightarrow \{0, 1\}$ definida por: $x \Delta_i y \Leftrightarrow d(x, y) \leq i \Leftrightarrow \min \{j \mid (x, y) \in P_j\} \leq i \Leftrightarrow (x, y) \in P_i$.

Bibliografia.

Jardine N. & Sibson R. (1971)
Mathematical Taxonomy
John Wiley & Sons Ltd.

Kendal, David G. (1969)
Some problems and methods in statisti-
cal archaeology
World Archaeology 1, 68-76

Kendal, David G., F.R.S. (1970)
A mathematical approach to
seriation.
Phil. Trans. Roy. Soc. Lond. A 269,
125-135.

Lerman, I.C. (1970)
Les bases de la classification
automatique
Collection Programmation
Paris: Gauthier-Villars

Sokal, R. R. & Sneath, P. H. A (1963)
Principles of numerical taxonomy
London & San Francisco: Freeman.