

**UNIVERSIDAD
DE
GRANADA**



**FACULTAD
DE
CIENCIAS**

**DEPARTAMENTO DE
ESTADISTICA E INVESTIGACION OPERATIVA**

**UNA APLICACION DE LA ENTROPIA AL ANALISIS
CLUSTER MEDIANTE VARIABLES CUALITATIVAS
MULTIESTADO: AFINIDAD, SIMILARIDAD Y
AGRUPACION**

JOSE ANTONIO ESQUIVEL GUERRERO

**TESIS DOCTORAL
GRANADA 1987**



Biblioteca Universitaria de Granada



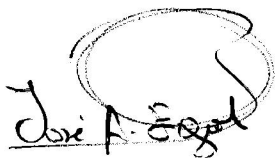
01179817

UNIVERSIDAD DE GRANADA. FACULTAD DE CIENCIAS


UNIVERSIDAD DE GRANADA
Facultad de Ciencias
23 SET. 1987
ENTRADA N.º 3806

UNA APLICACION DE LA ENTROPIA AL ANALISIS CLUSTER
MEDIANTE VARIABLES CUALITATIVAS MULTIESTADO: AFINIDAD,
SIMILARIDAD Y AGRUPACION.

BIBLIOTECA UNIVERSITARIA
GRANADA
Nº Documento <u>6 15514742</u>
Nº Copia <u>L 15824069</u>


José A. Esquivel

Tesis Doctoral presentada
por José Antonio Esquivel Gue-
rrero para la obtención del gra-
do de Doctor. Dirigida por Don
Ramón Gutiérrez Jáimez, Catedrá-
tico de Estadística e Investi-
gación Operativa de la Univer-
sidad de Granada.


Fdo. Ramón Gutiérrez Jáimez

Esta Tesis Doctoral no se habría podido realizar sin la ayuda y apoyo de varios compañeros de distintos Centros de investigación y Departamentos de la Universidad de Granada. En particular, los compañeros de los Departamentos de Estadística e Investigación Operativa y / de Prehistoria, por su ánimo constante en apoyo de este trabajo.

Hay que citar a otras personas por su ayuda en proveer, de forma/ desinteresada, los medios informáticos necesarios para el desarrollo/ de este trabajo. Fundamentalmente a E. Barahona de la Estación Experimental del Zaidín (CSIC), cuya cesión del ordenador del Centro (HP -/ 9816) permitió realizar los primeros bosquejos de los programas, y a V. Costa del Instituto de Astrofísica de Andalucía (CSIC) que permitió el trabajo en el ordenador del Centro (VAX 750), colaborando amplia y eficazmente en el manejo y ejecución de los programas.

Mención especial merece F. Contreras Cortés, del Departamento de/ Prehistoria, con quién ha sido posible el trabajo en equipo desde el/ comienzo de esta Tesis Doctoral y a quién se deben ideas e interpretaciones de tipo arqueológico importantes en el desarrollo de este trabajo.

Un profundo reconocimiento se debe a F. Molina González, Director del Departamento de Prehistoria, por el impulso y apoyo prestados a / la introducción de las técnicas de Análisis de Datos en esta disciplina. Ha proporcionado un apoyo esencial al comienzo de esta Teis, siendo de gran importancia sus ideas y sugerencias desde el punto de vista del prehistoriador.

Consideración especial y sincero agradecimiento merece el profesor R. Gutiérrez Jaimez, Director de esta Tesis Doctoral y del Departamento de Estadística e Investigación Operativa quién, desde el primer momento, ayudó y alentó el desarrollo de este trabajo con un apoyo y orientación constante. Sin su ayuda, ideas y dirección continua/ esta Tesis Doctoral no sería.

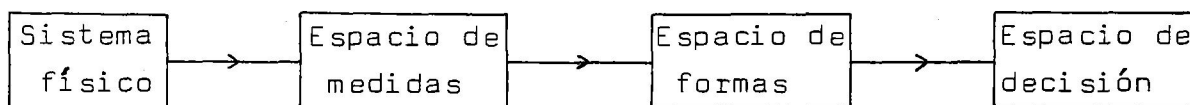
INDICE

Introducción	1
I.- Información y entropía de una variable	10
1.- Información de una variable	11
2.- Entropía	14
3.- Distorsión	20
II.- Perturbación de unidades	30
1.- Definiciones previas	30
2.- Teoremas del campo de una unidad	32
III.- Intersección y unión. Grupos	42
1.- Intersección	42
2.- Unión	44
IV.- Afinidad y similaridad	56
1.- Afinidad	56
2.- Afinidad grupo-grupo	66
3.- Similaridad inter-grupos	73
V.- Algoritmos de agrupación. Análisis Cluster	79
1.- La entropía de Shannon y el clustering	79
2.- La distorsión y la agrupación	81
3.- Algoritmo de clustering	88

Ejemplos 94
 Ejemplo I 94
 Ejemplo II 109
Apéndices 117
Conclusiones 122
Bibliografía 124

INTRODUCCION

El término Pattern Recognition constituye actualmente un amplio / campo de investigación que comprende, entre otros objetivos, la obtención de información, agrupación y clasificación a partir de sistemas/ físicos numéricos ó no-numéricos en un intento de estudiar los meca-/ nismos de reconocimiento de patrones utilizados por los humanos y los organismos vivientes (Arqueología, Biología, Psicología, etc.) y de / desarrollar las teorías y técnicas que permitan realizar este trabajo automaticamente (Ciencias de la Información, Computación, etc.) me- / diante un esquema de reconocimiento:



El objetivo de las técnicas de agrupación (clustering) ó clasifi- cación no supervisada es, fundamentalmente, encontrar una partición / de un conjunto de objetos en grupos de objetos similares de acuerdo a alguna medida de similaridad ó distancia (ó cualquier otro método) en tre objetos. Una vez realizada la agrupación, los resultados se utili- zan, generalmente, en dos vertientes:

- construcción de clases de patrones constituidas por objetos si- milares y claramente distinguidos unos de otros.
- técnica preliminar de análisis de datos en un intento de redu- cir el tamaño de los datos reteniendo la máxima información re- levante.

El proceso de agrupación intenta conseguir, en base a alguna medi- da de similaridad (distancia es equivalente ya que son recíprocas), / que los cluster (grupos) sean lo más homogéneos posible de forma que/ la similaridad entre objetos de un cluster sea mayor que la existente entre objetos pertenecientes a distintos clusters. Además, los clus- ters deben estar lo más aislados posible entre ellos, proporcionando/ una distinción clara entre objetos pertenecientes a distintos clus- /

ters.

Un problema fundamental aparece en la elección de la medida de similaridad como base de la agrupación, es decir, ¿Qué tipo de medida / de similaridad debe usarse para agrupar objetos? y ¿Qué algoritmos / hay que utilizar? (Michalski y Stepp (13), pp.396). Un estudio completo de estas técnicas puede encontrarse en Sneath y Sokal (15) entre otros.

Las técnicas de agrupación referidas a variables cuantitativas / han sido bastante desarrolladas y existen gran cantidad de medidas, / métodos de agrupación, algoritmos, técnicas de reagrupamiento, etc. / (Sneath y Sokal (15), Duda y Hart (19), Diday y Simon (18) y Escude-/ro (16)), incluso detección de clustering espacio-tiempo (McAuliffe y Afifi (25)). Sin embargo, la agrupación de un conjunto de datos definido mediante variables cualitativas comporta más dificultades producidas por el carácter de las variables que caracterizan a los objetos./ Los modelos más comunes están codificados mediante variables del tipo presencia-ausencia (1-0, etc.) en tablas 2x2 que estudian los acuer-/dos entre los estados (el uso de tablas 2x2 no indica necesariamente/ que los datos sean apropiados para un estudio mediante tablas de con-/tingencia) y existen gran cantidad de algoritmos y coeficientes de similaridad: coeficiente de Jacquard-Sneath, coeficiente métrico de Wi-/lliams y Lance, coeficiente de Sokal y Michener y corrección de Goo-/dall, coeficiente de Rogers-Tanimoto, coeficiente de Yule, etc. (en / (15) aparece una revisión completa de estos coeficientes).

Las variables cualitativas multiestado originan aún más problemas por el carácter de las variables y, con frecuencia, se han estudiado/ mediante técnicas de variables dicotómicas, considerando los estados/ de cada variable como un conjunto de variables dicotómicas (Kendall / (11), pp. 43-45, 171-198). El interés del problema en estudios y tra-/bajos de Pattern Recognition aplicados a campos como reconocimiento / de imágenes, reconocimiento de strings de símbolos, secuencias de fo-/nemas, estudio de documentos, estudio de unidades arqueológicas, orga-/nización de bases de datos, etc. ha suscitado el que sea abordado desde diversas teorías mediante técnicas no-paramétricas.

Desde las teorías fuzzy, las variables multiestado han originado / el desarrollo de índices de similaridad, técnicas de clustering, algo

ritmos, etc., fundamentalmente a partir de los trabajos de Ruspini / ((12), pp. 23-32) en los que se asigna a cada punto un cierto grado / de pertenencia a cada una de las clases S_1, \dots, S_n , introduciendo funciones de densidad para grupos $P(S_j)$ para conseguir minimizar (en media) los productos $P(S_j/X_i) \cdot M_j(x_i)$ de forma que grandes medias correspondan a pequeños grados de pertenencia, y viceversa. Otros intentos/ utilizan medidas de asociación entre un punto x y un conjunto fuzzy F como la inversa de la distancia media entre x y F , construyendo un algoritmo fuzzy basado en el clásico ISODATA. Backer introdujo un modelo de asociación que intenta encontrar una descomposición óptima del conjunto de objetos en una colección de conjuntos fuzzy mediante una medida de afinidad basada en las propiedades estructurales entre objetos, estudiando funciones de pertenencia a partir de distancias, del concepto de proximidad y de conceptos probabilísticos (Pal y Majumder (4), pp. 78-97, Backer y Jain (9), pp. 66-74 y Dubois y Prade (2), / pp. 323-331).

Recientemente, varias técnicas han abordado el problema desde distintas teorías: en base a la construcción de clases que representan / conceptos descriptivos mediante relaciones que envuelven afirmaciones conjuntivas y disyuntivas en el marco del cluster conceptual (Michalski y Stepp (13)), aún cuando existen opiniones que minimizan su importancia (Dale (20)); desde la teoría de grafos utilizando métodos de agrupación basados en el número de conexiones entre puntos de un grafo sin lazos y medidas de compacidad y separación que deben ser maximizadas (ambas ó una combinación lineal de ellas) (Bailey y Cowles (14), / pp. 645-649); generalizando medidas de Teoría de Conjuntos a variables no-independientes con aplicación al reconocimiento de patrones / de documentos en el campo del análisis del lenguaje (Ito, Kodama y Toyoda (5), 111-115); en base a caracteres jerárquicamente independientes, construyendo una jerarquía multinivelar previa de caracteres clasificados en primarios, secundarios, etc. para estudiar una medida de similaridad entre objetos arqueológicos no comparables ordinariamente (Ben-Bassat y Zaidenberg (17)); ó utilizando distancias fundadas en / las propiedades geométricas de los datos para estudiar objetos muy / disgregados (Tao Gu y Duibuisson (22), pp. 366-369).

La Teoría de la Información ha contribuido, casi desde sus orígenes, al estudio de las variables multiestado en áreas diversas que a-

barcan un amplio espectro: similaridad, clustering, entropía fuzzy, / discriminación, etc. A partir de la primitiva formulación de la dis- / tancia $d_{jk} = - \lg S_{jk}$ y la información de una unidad definidas por Rogers y Tanimoto, Goodall estableció un índice de similaridad computan- / do la entropía en base a las ocurrencias de los estados sobre las uni- / dades. Posteriormente, Estabrook y Orlocci ampliaron estas ideas con / las nociones de información conjunta e información mutua (detalles en / el Capítulo V) a partir de la definición de entropía de Shannon. Más / recientemente, la Teoría de la Información ha sido utilizada en cam- / pos diversos que abarcan:

- función de información-discriminación entre distribuciones de / probabilidad definidas sobre el mismo espacio, mediante tablas / de contingencia que conducen a modelos logaritmo-lineales / (Kullback y Keegel (23), pp. 134-137).
- problemas de decisión en un árbol clasificador (usando la en- / tropía de Shannon por su carácter aditivo). El proceso sigue / la técnica de reducir la entropía en cada nivel del árbol con / aplicación en agrupación, selección de patrones y diseño de de- /cisiones, utilizados en el reconocimiento de caracteres chinos / (Wang y Suen (3), pp. 406-413).
- evaluación de patrones mediante entropía fuzzy, definiendo un / índice de evaluación en términos de medidas inter e intracla- /ses para determinar y discriminar entre las diferentes clases / (Pal y Chakraborty (8), pp. 755-756).
- información total producida por conjuntos fuzzy en términos de / entropía fuzzy (con significado diferente a la entropía de Sha- /nnon), que equivale a la información media de Shannon produci- /da por N fuentes binarias de información (Xie y Bedrosian (7), / pp. 151-153 y Pal (21), pp. 205-206).

Esta Tesis Doctoral surge ante la necesidad de desarrollar estu- / dios tipológicos en Arqueología (sobre todo en Prehistoria) mediante / la aplicación de técnicas matemáticas a la clasificación de objetos / arqueológicos definidos por variables cualitativas multiestado. En es

ta disciplina son fundamentales los trabajos de Doran y Hodson (24) y Kendall (además de otros autores arqueólogos) que desarrollan métodos fundamentalmente de tipo cuantitativo en Arqueología.

El punto de partida lo constituye el intento de definir medidas / de afinidad y similitud que permitan construir un algoritmo de agrupación de unidades definidas mediante variables cualitativas multies-tado, en base a métodos y técnicas de Teoría de la Información. La idea básica se fundamenta en que, en el espacio de unidades, cada unidad crea una perturbación que afecta al resto de las unidades en función de la frecuencia de sus estados, y la perturbación que cada unidad produce sobre las demás permitirá encontrar una medida de la atracción que ejercen entre sí las unidades, grupos, etc.

En el Capítulo I se establecen las bases teóricas para definir una medida de información que tenga en cuenta la frecuencia de los estados, la incertidumbre que proporcionan y el número de estados de cada variable, lo que lleva a definir una entropía H_{n_i} de variables que verifica la axiomática exigida para las medidas de incertidumbre (Reza (2), pp. 82), estudiando las propiedades de H_{n_i} , tanto individualmente como en relación con la función H de Shannon. Posteriormente se define la función distorsión D_{n_i} y su ajuste a la axiomática exigida a las medidas de incertidumbre estudiándose las distintas propiedades y Teoremas que cuantifican las consecuencias de la subdivisión de un estado de una variable en varios estados, la agrupación de varios estados de una variable en uno sólo y la agrupación de varias variables en una única variable media, lo que permite predecir los resultados / de ampliar ó disminuir (en busca de una mayor información) los estados y/o las variables que definen los datos.

El Capítulo II está dedicado al estudio de la perturbación que / produce una unidad en el espacio de unidades en función de los estados que la conforman. El capítulo comienza estableciendo la noción de espacio de patrones como conjunto de objetos matemáticos que corresponde a una distribución de unidades y las probabilidades de cada estado en cada variable para extender la idea de distorsión de variables a la perturbación (denominada campo, en analogía con los campos gravitatorios) originada por un objeto en función de la información / que contiene. El espacio de patrones se entiende como un conjunto de /

fuerzas en equilibrio dinámico, equilibrio que se ve perturbado por / la introducción ó exclusión de una unidad ya que estos hechos afectan a la distribución de probabilidades de los estados. Después de estu- / diarse algunas propiedades del campo de una unidad, surgen una serie / de Teoremas y Corolarios que establecen las modificaciones producidas en el campo de una unidad en los casos:

- modificación (aumento ó disminución) del número de estados de / una variable sin modificar la probabilidad del estado que po- / see la unidad A, estableciendo límites a la modificación produ- / cida en el campo de A.
- modificación de la frecuencia (probabilidad) de los estados de / una variable sin variar el número de estados de la misma, estu- / diando el factor de variación en el campo de A y sus límites.
- modificaciones simultáneas del número y probabilidad de los es- / tados de una variable y la influencia que producen en el campo / de A.

La interacción entre unidades es un factor determinado por la es- / tructura del espacio y de las probabilidades de los estados. En el ca- / pítulo III se establece un álgebra en el conjunto de unidades median- / te operaciones conjuntivas y disjuntivas (denominadas, obviamente, in- / tersección y unión). La intersección de varias unidades engloba los / estados comunes a todas ellas, lo que permite definir la información / aportada por la intersección extendiendo la definición de campo de u- / na unidad al campo del objeto matemático intersección, además de estu- / diar las propiedades del mismo en función de los campos de las unida- / des individuales. Un proceso paralelo conduce a definir la unión de u- / nidades como el objeto matemático que sumariza la totalidad de los es- / tados que definen a las unidades, estableciendo el campo conjunto de / varias unidades que contiene la información global aportada por las / mismas y varias propiedades que relacionan el campo conjunto con los / campos de las unidades y el campo de la intersección. Dos Teoremas / fundamentales determinan las relaciones exactas entre el campo unión, / los campos de las unidades y el de la intersección, permitiendo calcu- / lar unos en función de los demás. Estos Teoremas conducen al cómputo / de las propiedades que relacionan los campos de unidades, interseccio- / nes y uniones múltiples como base para establecer el concepto de cam-

po asociado a un grupo de unidades, pieza fundamental en los conceptos de afinidad y similaridad entre grupos.

El punto de partida del Capítulo IV lo constituye el hecho de que el campo de un grupo contiene información tanto de la similitud entre sus elementos como de su disimilitud. El Capítulo comienza definiendo la afinidad elemento-grupo como extensión del campo intersección de / dos unidades de forma que compute la atracción entre el elemento y / los miembros del grupo para permitir reflejar las propiedades estructurales del mismo (Pal y Majumder (4), pp. 87). Un teorema fundamental expresa la relación entre la afinidad elemento-grupo y las intersecciones múltiples entre ellos, estudiando diversas propiedades respecto a las notas de la afinidad y la relación existente entre ésta y los campos unión elemento-grupo, del elemento y del grupo. Sucesivos Teoremas y Corolarios tratan los casos en que el elemento A pertenezca al grupo, estableciendo las condiciones en que la afinidad alcanza un máximo absoluto. Posteriormente se define la afinidad entre un grupo y la intersección de dos elementos, previo estudio de la invariancia al permutar un elemento de un grupo a otro para dar cohesión a la definición. Esta afinidad elemento-grupo verifica las propiedades intuitivas establecidas por Backer y Jain, y difundidas por Pal y Majumder, que debe cumplir cualquier medida de afinidad.

De forma natural, la afinidad elemento-grupo se extiende a afinidad grupo-grupo manteniendo la información producida por las intersecciones múltiples de los grupos. Sus propiedades quedan establecidas / con referencia a los campos de los grupos, de la intersección, las relaciones de inclusión entre los grupos y, fundamentalmente, la relación existente entre afinidad, campos de grupos y el campo conjunto / de ambos, base indispensable para un proceso de agrupación en el que los grupos se fusionen en uno sólo. Posteriormente queda establecida / la extensión de las exigencias de Pal y Majumder a la afinidad grupo-grupo y se estudia su cumplimiento por la definición anterior, además de definir un método de cálculo para la afinidad que consuma menos operaciones, tiempo y complejidad como alternativa al cálculo de las / intersecciones múltiples entre todos los elementos de los grupos.

Las peculiaridades de la afinidad inducen a definir unas medidas / de similaridad que tengan en cuenta no sólo la atracción numérica en-

tre ellos sino también toda la información aportada por los grupos, / creando dos medidas básicas relativas al campo de los grupos para con- seguir la comparación efectiva de las similitudes entre grupos sin/ las interferencias que producen los valores muy altos ó muy bajos en/ la afinidad debidos a acuerdos ó desacuerdos en estados poco ó muy / frecuentes. Basicamente se establecen dos medidas de similitud:

- similitud fuerte, basada en la intersección fuerte entre los grupos (intersección de todos los elementos de ambos grupos) / que computa la información, relativa al campo unión, proporció- nada por el núcleo que cohesionaría a los grupos si se fusiona- ran en uno sólo. Esta similitud verifica la extensión de las propiedades de Pal y Majumder y se muestra adecuada para reali- zar una agrupación jerárquica de acuerdo a las técnicas SAHN.
- similitud basada en la afinidad, que computa todas las cone- xiones entre los elementos de los grupos, estableciendose su a- decuación a las propiedades de Pal y Majumder. Esta medida ex- hibe algún inconveniente para realizar una agrupación jerárqui- ca, pero puede subsanarse adecuadamente realizando una modifi- cación en la jerarquía.

El Capítulo V y último está dedicado al clustering y algoritmos / de agrupación, con un bosquejo previo de la relación existente entre/ la entropía de Shannon y el análisis cluster. Se establecen las rela- ciones entre la afinidad y la agrupación estudiada desde el punto de/ vista del análisis cluster, con especial atención a la similitud / fuerte y a la similitud-afinidad de acuerdo a las características / propias de cada una de ellas y en función de sus valores extremos, re- lativización al campo conjunto y decrecimiento con el número de ele- mentos en los grupos una vez realizada la agrupación. Varias medidas/ derivadas de las anteriores son propuestas en un intento de moderar / algunos efectos negativos producidos por los grupos con muchos miem- / bros.

Los algoritmos de agrupación ocupan el resto del Capítulo y son / expuestos con detalle, al principio como un todo completo y posterior- mente describiendo los procedimientos fundamentales CLAGRUP, SIMI, / LOOP y RECAL que realizan los cálculos básicos para conseguir la agru- pación según:

- CALGRUP localiza los líderes de cada grupo y calcula la similitud entre todos los grupos.
- SIMI marca los elementos de los grupos, calcula la similaridad grupo-grupo y el campo del grupo fusión (en el caso de que se realice).
- LOOP ciclo de agrupación computando la mayor similaridad, nivel y final de la agrupación, etc.
- RECAL actualiza los parámetros necesarios una vez realizada la agrupación en cada nivel.

Otros procedimientos auxiliares completan el algoritmo. El procedimiento AFIN calcula la afinidad fuerte (intersección) permitiendo que, junto a los anteriores, el algoritmo trabaje con los dos principales índices de similaridad S_1 y S_2 . Los procedimientos auxiliares IGUAL (determina si dos unidades coinciden ó no en todos sus estados) y NUMEL (computa los elementos repetidos de cada futuro grupo) posibilitan la adaptación del algoritmo para su utilización con las medidas derivadas de la similaridad fuerte y la similaridad-afinidad.

CAPITULO I

El estudio de una distribución de unidades con variables cualitativas multiestado exige obtener la máxima información de cada unidad, de cada variable y de cada estado en función de la frecuencia de aparición, tipo de estado, etc. Términos iguales a rareza, abundancia y/o otras nociones intuitivas deben reflejarse en el estudio de una forma clara y precisa.

De acuerdo a las teorías de Shannon, en un modelo matemático de comunicación la información vendrá determinada por un parámetro estadístico asociado a un esquema de probabilidad y "el parámetro debe indicar una medida relativa a la incertidumbre de acuerdo a la ocurrencia de cada mensaje particular en el conjunto de mensajes" (sic) (Rezza (1), pp. 6). Estas ideas van a constituir el núcleo del análisis, aunque teniendo en cuenta las lógicas diferencias existentes entre una distribución de unidades y la información que proporciona un determinado mensaje.

En el "espacio" de unidades, cada unidad produce una perturbación que viene determinada por su propia incertidumbre de acuerdo a la incertidumbre que le proporciona cada uno de sus estados en cada variable. El "espacio" de unidades se ve distorsionado con la perturbación que le aporta cada unidad, de forma que una unidad "rara" creará una distorsión grande en el espacio y una unidad que posea estados muy abundantes en la distribución general producirá poca distorsión en la homogeneidad del espacio. Por tanto, la función de evaluación de la entropía o "distorsión" verificará:

- I) Cada estado aporta un peso específico a la distorsión en función de la frecuencia del mismo y del número de estados de la variable a la que pertenece.
- II) La incertidumbre de un estado muy frecuente debe ser pequeña puesto que la probabilidad de que aparezca en una unidad escogida al azar es grande; reciprocamente, si un estado es raro, su contribución a la entropía debe ser grande.
- III) La perturbación que introduce una variable en el espacio será mayor cuanto menor sea el número de sus estados ya que /

una variable con pocos estados produce una dicotomía mayor en la distribución de unidades en el espacio que otra con numerosos estados.

1. INFORMACION DE UNA VARIABLE.

Sea una distribución de unidades con valores en el conjunto de variables multiestado $V = \{X_1, X_2, \dots, X_v\}$ tal que cada variable X_i tiene asociado un conjunto de estados según:

$$\begin{aligned} X_1 &\longrightarrow W_1 = \{x_{11}, x_{12}, \dots, x_{1n_1}\} \\ X_2 &\longrightarrow W_2 = \{x_{21}, x_{22}, \dots, x_{2n_2}\} \\ &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ X_i &\longrightarrow W_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\} \\ &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ X_v &\longrightarrow W_v = \{x_{v1}, x_{v2}, \dots, x_{vn_v}\} \end{aligned}$$

siendo $N(V)=v$ (número de variables) y $N(W_i)=n_i$, $i=1,2,\dots,v$ (número / de estados en la variable X_i).

Una unidad vendrá determinada por los valores particulares que toma cada variable en la unidad de forma que cada variable esté representada por uno y sólo uno de sus estados en cada unidad. La representación de la unidad a_j será de la forma

$$a_j = \{a_j^1, a_j^2, \dots, a_j^v\}$$

siendo $a_j^k = x_{kh}$, $1 \leq h \leq n_k$ y $1 \leq k \leq v$, si a_j toma el estado h en la variable k , es decir

$$a_j^k \in W_k \quad \text{y} \quad a_j \in W_1 \times W_2 \times \dots \times W_v$$

Naturalmente, si el máximo número de unidades es n , puede ocurrir que sea $n > n_1 \cdot n_2 \cdot \dots \cdot n_v$ ya que pueden existir varias unidades que tomen los mismos valores en una, varias o todas las variables.

1.1 Frecuencias.

Cada estado posee una frecuencia de aparición en la distribución/

de unidades en función del número de unidades que lo posee

$$f(x_{jk}) = \frac{n_{jk}}{n}, \quad 0 < f(x_{jk}) \leq 1,$$

siendo n el número de unidades en la distribución.

Cada variable tiene asociado un espacio de frecuencias en función de la frecuencia de cada uno de sus estados

$$W_i \longrightarrow P(W_i) = \{f(x_{i1}), f(x_{i2}), \dots, f(x_{in_i})\},$$

$i=1, 2, \dots, v$. Evidentemente se tiene que

$$\sum_{k=1}^{n_i} f(x_{ik}) = 1, \quad i=1, 2, \dots, v$$

En el caso de una distribución de tipo probabilístico las frecuencias aparecerán como probabilidades (a veces se utilizará dicha notación).

1.2 Información

Considerando la variable aleatoria $X = -\lg p(x)$, a cada estado x_{ik} le corresponde un valor X_{ik} en la variable aleatoria X de acuerdo a

$$X_{ik} = -\lg p(x_{ik}) \quad (X_{ik} = -\lg f(x_{ik}) \text{ según frecuencias}).$$

La cantidad de información del estado x_{ik} se define entonces como

$$I(x_{ik}) = -\frac{1}{n_i} \cdot \lg p(x_{ik}) \geq 0$$

La información correspondiente a un estado solamente será 0 si el estado no aparece en ninguna unidad o si es el único estado de la variable, en cuyo caso la variable no aporta ninguna información ya que todas las unidades poseerían el mismo estado en dicha variable.

El logaritmo se toma en base 2 de acuerdo al formulismo de la Teoría de la Información aunque no habría inconveniente en elegir otra base. Esta definición de información se ajusta a la axiomática de Shannon ya que cualquier función de la forma $k \cdot \lg(x)$ es una solución de la ecuación

$$f\left(\frac{1}{n}\right) + f\left(\frac{1}{m}\right) = f\left(\frac{1}{mn}\right).$$

Este parámetro constituye una medida de la perturbación que introduce el estado x_{ik} en el espacio a través de las unidades que lo poseen, afectando a la homogeneidad de la distribución en mayor o menor medida.

La información definida anteriormente verifica de forma trivial /

las exigencias I), II) y III) puesto que:

I) por construcción.

II) dados $x_{ik}, x_{ij} \in W_i$, si $p(x_{ik}) < p(x_{ij})$ entonces se tiene / que $\lg p(x_{ik}) < \lg p(x_{ij})$ al ser \lg creciente. De aquí se deduce que $-\lg p(x_{ik}) > -\lg p(x_{ij})$ y $-\frac{1}{n_i} \lg p(x_{ik}) > / > -\frac{1}{n_i} \lg p(x_{ij})$, por lo que $I(x_{ik}) > I(x_{ij})$.

La discriminación que proporciona un estado poco frecuente mediante la información que aporta es mayor que la producida por otro estado más común.

III) sean $x_{ik} \in W_i$ y $x_{jp} \in W_j$ tales que $p(x_{ij}) = p(x_{jp})$ y que $n_i < n_j$. Entonces $-\frac{1}{n_i} < -\frac{1}{n_j}$ y como $\lg p(x) < 0$ al ser $p(x) < 1$, se tiene que $-\frac{1}{n_i} \lg p(x_{ik}) > -\frac{1}{n_j} \lg p(x_{jp})$ y de aquí: $I(x_{ik}) > I(x_{jp})$.

Este resultado indica que si dos estados son equiprobables la mayor información la proporciona el estado que pertenezca a la variable con menor número de estados.

1.3 Propiedad.

Sean los estados $x_{ik} \in W_i$ y $x_{jp} \in W_j$. Entonces, si $n_i < n_j$ se tiene:

a) si $p(x_{ik}) < p(x_{jp})$ entonces $I(x_{ik}) > I(x_{jp})$.

b) si $p(x_{ik}) > p(x_{jp})$ entonces $I(x_{ik}) < I(x_{jp})$.

a) por el crecimiento de $\lg x$, $-\lg p(x_{ik}) > -\lg p(x_{jp})$, y / como $\frac{1}{n_i} > \frac{1}{n_j} > 0$ se obtiene $-\frac{1}{n_i} \lg p(x_{ik}) > / > -\frac{1}{n_j} \lg p(x_{jp})$, es decir, $I(x_{ik}) > I(x_{jp})$.

b) de forma simétrica se deduce que $\lg p(x_{ik}) > \lg p(x_{jp})$ y $\frac{1}{n_i} \lg p(x_{ik}) > \frac{1}{n_j} \lg p(x_{jp})$, de donde $I(x_{ik}) < I(x_{jp})$.

Esta propiedad pone de manifiesto que la influencia de la probabilidad en la información es mayor que la que produce el número de estada

dos. Además, la combinación probabilidad pequeña y escaso número de / estados produce la máxima información y, por tanto, la máxima contribución a la distorsión del espacio.

2. ENTROPIA.

La entropía es un concepto clásico dentro de la teoría de la información para medir la incertidumbre, definida como la media estadística de un esquema de probabilidad finito y completo $H(x) = \overline{I_k}$. Para cada variable X_i hay una definición de entropía según

$$H_{n_i}(X_i) = \overline{I(x_{ik})}, \quad k=1,2,\dots,n_i$$

De acuerdo a la definición de información 1.2, la función H_{n_i} está definida como

$$H_{n_i}(X_i) = H_{n_i}(p(x_{i1}), \dots, p(x_{in_i})) = - \frac{1}{n_i} \cdot \sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg p(x_{ik})$$

Este parámetro proporciona una medida de la contribución media de cada variable a la incertidumbre del espacio y, de acuerdo a su definición, tiene las siguientes propiedades de trivial demostración:

2.1 Propiedades elementales de H.

Sean las variables $X_i, X_j \in V$. Entonces:

$$1) \text{ si } \sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg p(x_{ik}) = \sum_{p=1}^{n_j} p(x_{jp}) \cdot \lg p(x_{jp}) \text{ y } /$$

$$n_i < n_j, \text{ entonces } H_{n_i}(X_i) > H_{n_j}(X_j).$$

2) si $n_i < n_j$ se verifica:

$$a) \text{ si } \sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg p(x_{ik}) < \sum_{p=1}^{n_j} p(x_{jp}) \cdot \lg p(x_{jp}) \text{ entonces}$$

$$\text{ces } H_{n_i}(X_i) > H_{n_j}(X_j).$$

b) si
$$\sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg p(x_{ik}) > \sum_{p=1}^{n_j} p(x_{jp}) \cdot \lg p(x_{jp})$$
 entonces $H_{n_i}(X_i) < H_{n_j}(X_j)$.

Existe una relación clara entre la entropía clásica de Shannon /

$$H(X_i) = - \sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg p(x_{ik})$$

y H_{n_i} que permite una comparación entre ambas entropías (definición / de Shannon-Wiener). Como $H_{n_i}(X_i) = \frac{1}{n_i} \cdot H(X_i)$, dos variables X_i y X_j / con igual entropía de Shannon ($H(X_i) = H(X_j)$) no tienen la misma entropía H_{n_i} puesto que el número de estados de cada variable es decisivo: a igual entropía H , la mayor entropía H_{n_i} corresponde a la variable con menor número de estados.

2.2 La función H_{n_i} como medida de incertidumbre.

La definición axiomática de una función como medida de incertidumbre debe verificar, de acuerdo a las teorías de Shannon (Rezza (1)), / las condiciones:

I) continuidad. $H(p_1, \dots, p_n)$ debe ser continua en p_k , $k=1, \dots, n$

II) simetría. $H(p_1, \dots, p_i, \dots, p_k, \dots, p_n) = H(\dots, p_k, \dots, p_i, \dots)$ / $1 \leq i \leq k \leq n$.

III) existencia de extremo. para H el máximo es $H(\frac{1}{n}, \dots, \frac{1}{n})$.

IV) aditividad. si $p_n = \sum_{k=1}^t q_k$ debe verificarse:

$$H(p_1, \dots, p_{n-1}, q_1, \dots, q_t) = H(p_1, \dots, p_n) + p_n H\left(\frac{q_1}{p_n}, \dots, \frac{q_t}{p_n}\right).$$

Varios autores (fundamentalmente Fadiev, Khinchin y Schutzenberger) han estudiado la función H cuando no verifica alguna de estas / propiedades surgiendo incluso otras definiciones de entropía (p.e. R.

A. Fisher). Sin embargo la más usada es la anterior definición de Shannon-Wiener debido a su simplicidad y a la verificación de la condición IV) (Wnag y Suen, (3) pp. 407).

La función H_{n_i} verifica, para cada X_i , unas condiciones equivalentes a I), II), III) y IV), con las modificaciones necesarias de acuerdo a las peculiaridades y diferencias conceptuales entre ambas definiciones, aunque existe cierto paralelismo entre la información que proporciona una unidad de una distribución y la producida por un mensaje del alfabeto:

I') $H_{n_i}(p(x_{i1}), \dots, p(x_{in_i}))$ es una función continua en cada una de las variables $p(x_{ik})$ en el intervalo $(0, 1]$, $k=1, \dots, n_i$ y $i=1, \dots, v$.

la demostración aparece en Rezza ((1), pp. 83) basada en la continuidad de $\lg x$.

II') la función H_{n_i} es simétrica en cada variable $p(x_{ik})$

por definición.

III') H_{n_i} tiene un máximo en $H_{n_i}(\frac{1}{n_i}, \dots, \frac{1}{n_i})$ para X_i con estados equiprobables ($p(x_{ik}) = \frac{1}{n_i}$, $k=1, \dots, n_i$).

siguiendo las líneas de la demostración III) que aparece en Rezza ((1), pp. 86-89) y la convexidad de $\log x$, se tiene

$$H_{n_i}(\frac{1}{n_i}, \dots, \frac{1}{n_i}) = - \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{1}{n_i} \cdot \lg \frac{1}{n_i} = \frac{\lg n_i}{n_i}, \text{ de donde /}$$

$$H_{n_i}(X_i) - \frac{\lg n_i}{n_i} = - \frac{1}{n_i} \left[\sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg p(x_{ik}) + \lg n_i \right] = /$$

$$= - \frac{1}{n_i} \left[\sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg \frac{1}{p(x_{ik})} + \lg \frac{1}{n_i} \right], \text{ y } \sum_{k=1}^{n_i} p(x_{ik}) = 1 \text{ de/}$$

$$\text{donde } H_{n_i}(X_i) - \frac{\lg n_i}{n_i} = \frac{1}{n_i} \left[\sum_{k=1}^{n_i} p(x_{ik}) \cdot \left(\lg \frac{1}{p(x_{ik})} + \lg \frac{1}{n_i} \right) \right]$$

$$= \frac{1}{n_i} \sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg \frac{1}{n_i p(x_{ik})} . \text{ Por la convexidad de } \lg x \text{ se tiene que}$$

$\ln x \leq x-1, x \in \mathbb{R}^+$ y el cambio de base es $\lg x = \ln x \cdot \lg_2 e$, de donde /

$$H_{n_i}(X_i) - \frac{\lg n_i}{n_i} \leq \frac{1}{n_i} \left[\sum_{k=1}^{n_i} p(x_{ik}) \left(\frac{1}{n_i p(x_{ik})} - 1 \right) \right] \cdot \lg_2 e = /$$

$$= \frac{1}{n_i} \left(n_i \frac{1}{n_i} - 1 \right) \cdot \lg_2 e = 0, \text{ esto es, } H_{n_i}(X_i) \leq \frac{\lg n_i}{n_i} \text{ y, por consiguient-}$$

te, $H_{n_i} \left(\frac{1}{n_i}, \dots, \frac{1}{n_i} \right)$ es un máximo para H_{n_i} .

IV') sea $H_{n_i+k-1}(p(x_{i1}), \dots, p(x_{in_i-1}), q_1, \dots, q_k)$ tal que verifi-

ca $\sum_{j=1}^k q_j = 1$. Entonces se verifica la relación:

$$(n_i+k-1)H_{n_i+k-1} = n_i H_{n_i} + k \cdot p(x_{in_i}) \cdot H_k \left(\frac{q_1}{p(x_{in_i})}, \dots, \frac{q_k}{p(x_{in_i})} \right) .$$

por definición de H_{n_i+k-1} se obtiene la expresión siguiente:

$$(n_i+k-1)H_{n_i+k-1}(p(x_{i1}), \dots, p(x_{in_i}), q_1, \dots, q_k) = - \left[\sum_{p=1}^{n_i-1} p(x_{ip}) \lg p(x_{ip}) \right.$$

$$\left. + \sum_{j=1}^k q_j \lg q_j \right] = - \left[\sum_{p=1}^{n_i-1} p(x_{ip}) \cdot \lg p(x_{ip}) + p(x_{in_i}) \cdot \lg p(x_{in_i}) - /$$

$$- p(x_{in_i}) \cdot \lg p(x_{in_i}) + \sum_{j=1}^k q_j \lg q_j \right] = - \left[\sum_{p=1}^{n_i} p(x_{ip}) \cdot \lg p(x_{ip}) - /$$

$$- \sum_{j=1}^k q_j \lg p(x_{in_i}) + \sum_{j=1}^k q_j \lg q_j \right] = /$$

$$\begin{aligned}
&= - \left[\sum_{p=1}^{n_i} p(x_{ip}) \cdot \lg p(x_{ip}) + \sum_{j=1}^k q_j \lg \frac{q_j}{p(x_{in_i})} \right] = / \\
&= - \left[n_i \frac{1}{n_i} \sum_{p=1}^{n_i} p(x_{ip}) \cdot \lg p(x_{ip}) + k \cdot p(x_{in_i}) \cdot \frac{1}{k} \sum_{j=1}^k \frac{q_j}{p(x_{in_i})} \cdot \lg \frac{q_j}{p(x_{in_i})} \right] = \\
&= n_i H_{n_i}(p(x_{i1}), \dots, p(x_{in_i})) + k \cdot p(x_{in_i}) \cdot H_k\left(\frac{q_1}{p(x_{in_i})}, \dots, \frac{q_k}{p(x_{in_i})}\right).
\end{aligned}$$

En forma abreviada, estas conclusiones pueden ponerse como

$$(n_i+k-1)H_{n_i+k-1} = n_i H_{n_i} + k \cdot p_{n_i} \cdot H_k$$

esto es:

$$H_{n_i+k-1} = \frac{n_i}{n_i+k-1} \cdot H_{n_i} + \frac{k}{n_i+k-1} \cdot p_{n_i} \cdot H_k.$$

Las diferencias entre IV y IV' se deben a la influebcia del término $1/n_i$ en la definición de H_{n_i} . La subdivisión de X_{in_i} en los estados q_1, \dots, q_k tiene como consecuencias que X_{in_i} se convierta en una "subvariable" con k estados que determina la diferencia entre ambas propiedades. Esta diferencia puede cuantificarse mediante un residuo $R_{n_i k}$, función de H_{n_i} y H_k , que parametriza la influencia introducida por los estados q_j :

puesto que $\frac{n_i}{n_i+k-1} = 1 - \frac{k-1}{n_i+k-1}$ y $\frac{k}{n_i+k-1} = 1 - \frac{n_i-1}{n_i+k-1}$, se tiene que

$$H_{n_i+k-1} = \left(1 - \frac{k-1}{n_i+k-1}\right) \cdot H_{n_i} + \left(1 - \frac{n_i-1}{n_i+k-1}\right) \cdot p_{in_i} \cdot H_k, \text{ es decir/}$$

$H_{n_i+k-1} = H_{n_i} + p_{in_i} \cdot H_k - \left[\frac{k-1}{n_i+k-1} \cdot H_{n_i} + \frac{n_i-1}{n_i+k-1} \cdot p_{in_i} \cdot H_k \right]$. Haciendo el valor del resto igual a la expresión entre corchetes del resultado anterior se tiene que:

$$R_{n_i k} = \frac{k-1}{n_i+k-1} \cdot H_{n_i} + \frac{n_i-1}{n_i+k-1} \cdot p_{i n_i} \cdot H_k \quad \text{equivalente a} \quad /$$

$$H_{n_i+k-1} = H_{n_i} + p_{i n_i} \cdot H_k - R_{n_i k} \quad \text{ó} \quad H_{n_i} + p_{i n_i} \cdot H_k = H_{n_i+k-1} + R_{n_i k}$$

2.3 Propiedades de la función H_{n_i} .

Estas propiedades relacionan $H_{n_i}(X_i)$ con la entropía H de Shannon en función del valor de n_i .

1) si $H(X_i) = H(X_j)$ y $n_i < n_j$, entonces $H_{n_i}(X_i) > H_{n_j}(X_j)$.

como $n_i < n_j$ entonces $\frac{1}{n_j} < \frac{1}{n_i}$, de donde $\frac{1}{n_j} H(X_j) < \frac{1}{n_i} H(X_i)$ /
es decir, $H_{n_i}(X_i) > H_{n_j}(X_j)$.

2) si $n_i < n_j$, entonces:

a) si $H(X_i) > H(X_j)$ se tiene que $H_{n_i}(X_i) > H_{n_j}(X_j)$.

como $\frac{1}{n_i} > \frac{1}{n_j}$ se obtiene que $H_{n_i}(X_i) > H_{n_j}(X_j)$.

b) si $H(X_i) < H(X_j)$ se tiene que $H_{n_i}(X_i) < H_{n_j}(X_j)$.

trivial mediante un razonamiento igual al de a).

3) para cada variable X_i , el aumento en el número de estados disminuye la incertidumbre máxima si $n_i \geq 3$.

por III'), el máximo para X_i es $\frac{\lg n_i}{n_i}$ y teniendo en cuenta que $\frac{\lg x}{x} \rightarrow 0$ si $x \rightarrow \infty$, a partir de un n_i en adelante /

la incertidumbre será menor cuanto mayor sea el número de estados. Y del estudio de la función $\frac{\lg x}{x}$ se deduce que tiene un/máximo en $x=e$ siendo creciente para $x < e$ y decreciente para $x > e$ por lo que el máximo es decreciente para $n_i \geq 3$.

3. DISTORION.

La entropía y la función H_{n_i} constituyen una medida de la información media que proporciona una variable y, por tanto, tienden a suavizar la influencia de los estados extremos (aquellos que tienen frecuencia muy pequeña o muy grande). Sin embargo, la influencia de estos estados es fundamental para el estudio de la distribución puesto que la coincidencia de dos unidades en un estado determinado debe valorarse en función de la información completa que aporta. Si esta coincidencia se produce en un estado poco frecuente su significación será mayor que si el estado es muy frecuente, de acuerdo con los coeficientes probabilísticos de similaridad que se basan en la filosofía "... el acuerdo entre estados raros es un suceso menos probable que el acuerdo entre estados frecuentes y debe ser valorado más ..." (Sneath y Sokal, (11) pp. 140-145). Además, el número de estados de la variable influirá negativamente en la afinidad entre unidades ya que es menos significativa una coincidencia en un estado de una variable con muchos estados que la coincidencia si los estados de la variable son escasos.

Estas consideraciones inducen la definición de la "distorsión" que produce una variable como:

3.1 Distorsión de una variable.

Sea la variable X_i con n_i estados. La distorsión de X_i se define según:

$$D_{n_i}(X_i) = \sum_{k=1}^{n_i} I(X_i) \quad \text{siendo} \quad \sum_{k=1}^{n_i} p(x_{ik}) = 1$$

que es equivalente a

$$D_{n_i}(X_i) = D_{n_i}(p(x_{i1}), \dots, p(x_{in_i})) = - \frac{1}{n_i} \sum_{k=1}^{n_i} \lg p(x_{ik}).$$

La distorsión de una variable refleja la influencia que produce / cada estado en el espacio de unidades y cómo queda afectada la homogeneidad de dicho espacio en función de la información que aporta cada/

una de las unidades.

A partir de la definición se deducen las propiedades siguientes: /

3.2 Propiedades de la distorsión D_{n_i} .

1) si $\prod_{k=1}^{n_i} p(x_{ik}) = \prod_{h=1}^{n_j} p(x_{jh})$ y $n_i < n_j$, entonces $D_{n_i}(X_i) > D_{n_j}(X_j)$

como $\lg(ab) = \lg a + \lg b$, aplicar lo a la igualdad anterior /

se tiene que $-\sum_{k=1}^{n_i} \lg p(x_{ik}) = -\sum_{h=1}^{n_j} \lg p(x_{jh})$, y $\frac{1}{n_i} > \frac{1}{n_j}$ al /

ser $n_i < n_j$, de donde $D_{n_i}(X_i) > D_{n_j}(X_j)$.

Esta propiedad refleja que, si las probabilidades se mantie- /
nen constantes en los términos antes expresados, la variable /
con menor número de estados produce mayor distorsión.

2) para cada $X_i \in V$ se verifica $H_{n_i}(X_i) < D_{n_i}(X_i)$.

como $\lg p(x_{ik}) < 0$ se tiene que $-\lg p(x_{ik}) > 0$ y, al ser /
 $p(x_{ik}) < 1$ resulta $-p(x_{ik}) \cdot \lg p(x_{ik}) < -\lg p(x_{ik})$ (multiplican /
do las desigualdades). Sumando todas las desigualdades en k se

obtiene $-\sum_{k=1}^{n_i} p(x_{ik}) \cdot \lg p(x_{ik}) < -\sum_{k=1}^{n_i} \lg p(x_{ik})$.

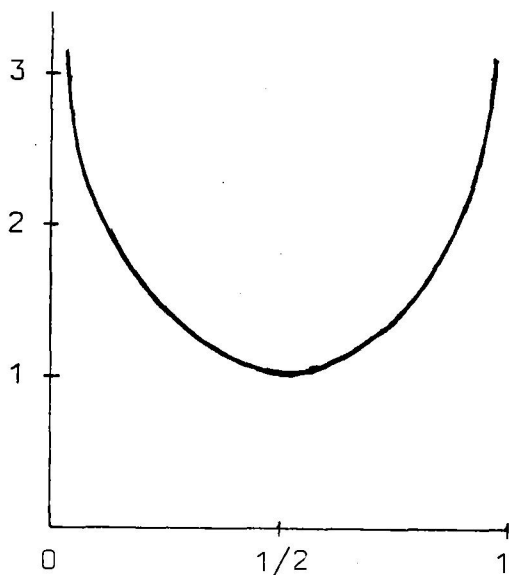
3) En el caso de una variable binaria se puede construir la grá- /
fica de la distorsión D_{n_i} y compararla con la gráfica de H:

en el caso de una variable binaria X se tiene que el núme- /
ro de estados es $n_i = 2$. Los estados constituyentes tienen las /
probabilidades

$$p(x_1) = x, \quad p(x_2) = 1-x$$

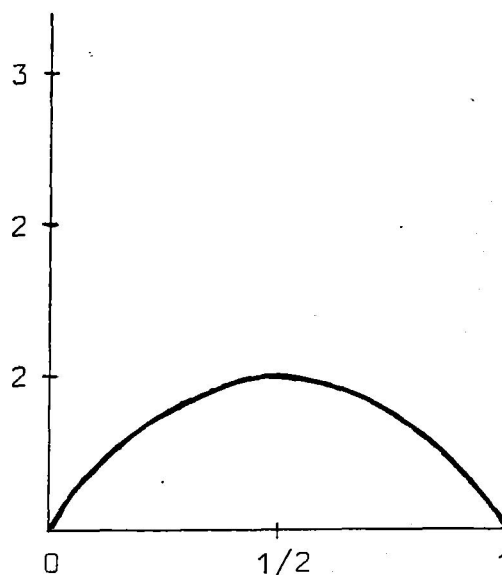
y las funciones entropía y distorsión toman la forma

$$D(X) = -\frac{1}{2}(\lg x + \lg(1-x)) \text{ y } H(X) = -\frac{1}{2}(x \lg x + (1-x) \lg(1-x))$$



gráfica $D(X)$

V
A
R
I
A
B
L
E
B
I
N
A
R
I
A



gráfica $H(X)$

3.3 La distorsión como medida de incertidumbre.

La definición de D_{n_i} como medida de información total que aporta una variable al espacio no se corresponde minuciosamente con la clásica entropía de Shannon definida como información media. Este hecho / significa que no se cumplen las propiedades I), II), III) y IV) en su totalidad sino una adaptación de las mismas que tengan en cuenta las características de D_{n_i} (I_D , II_D , III_D y IV_D), lo que permite comparar las analogías y diferencias entre ambos conceptos.

I_D) la función $D_{n_i}(p(x_{i1}), \dots, p(x_{in_i}))$ es continua en $p(x_{ik})$ en el intervalo $(0, 1]$, $k=1, \dots, n_i$, $i=1, \dots, v$.

la demostración es trivial siguiendo las líneas de la demostración de I') en base a la continuidad de $\lg x$.

II_D) la función D_{n_i} es simétrica en cada variable.

por definición.

III_D) D_{n_i} alcanza un mínimo cuando los estados de X_i son equiprobables $p(x_{ik}) = \frac{1}{n_i}$, $k=1, \dots, n_i$.

la demostración sigue las líneas de Rezza ((1), pp. 86-89) en base a que $\ln x \leq x-1$ si $x \in \mathbb{R}^+$ al ser $\lg x$ convexa:

$$D_{n_i} \left(\frac{1}{n_i}, \dots, \frac{1}{n_i} \right) = - \frac{1}{n_i} \sum_{k=1}^{n_i} \lg \frac{1}{n_i} = \lg n_i. \text{ Entonces } /$$

$$\lg n_i - D_{n_i}(X_i) = \lg n_i + \frac{1}{n_i} \sum_{k=1}^{n_i} \lg p(x_{ik}) = \frac{1}{n_i} \left[n_i \lg n_i + \sum_{k=1}^{n_i} \lg p(x_{ik}) \right] = /$$

$$= \frac{1}{n_i} \sum_{k=1}^{n_i} \lg [p(x_{ik}) \cdot n_i]. \text{ Al ser } \lg x = \ln x \cdot \lg_2 e \text{ y } \ln x \leq x-1, x \in \mathbb{R}^+ \text{ se } /$$

$$\text{tiene que } \lg n_i - D_{n_i}(X_i) = \frac{\lg_2 e}{n_i} \sum_{k=1}^{n_i} \ln [p(x_{ik}) \cdot n_i], \text{ de donde } /$$

$$\lg n_i - D_{n_i}(X_i) \leq \frac{\lg_2 e}{n_i} \sum_{k=1}^{n_i} (p(x_{ik}) \cdot n_i - 1) = \frac{\lg_2 e}{n_i} \cdot \left[n_i \sum_{k=1}^{n_i} p(x_{ik}) - n_i \right] = 0 /$$

al ser $\sum_{k=1}^{n_i} p(x_{ik}) = 1$. Entonces $D_{n_i}(X_i) \geq \lg n_i$ y el mínimo para D_{n_i} es/

$$D_{n_i}(X_i) = D_{n_i} \left(\frac{1}{n_i}, \dots, \frac{1}{n_i} \right) = \lg n_i.$$

IV_D) sean q_1, \dots, q_k tales que verifican $\sum_{k=1}^k q_j = p(x_{in_i})$. Entonces

$$D_{n_i+k-1}(p(x_{i1}), \dots, p(x_{in_i-1}), q_1, \dots, q_k) = \frac{n_i}{n_i+k-1} \cdot D_{n_i}(p(x_{i1}), \dots, p(x_{in_i})) +$$

$$+ \frac{k}{n_i+k-1} \cdot D_k \left(\frac{q_1}{p(x_{in_i})}, \dots, \frac{q_k}{p(x_{in_i})} \right) + \frac{k-1}{n_i+k-1} \cdot p(x_{in_i}).$$

$$\begin{aligned}
& (n_{i+k-1}) D_{n_{i+k-1}}(p(x_{i1}), \dots, p(x_{in_i}), q_1, \dots, q_k) = / \\
& = - \frac{1}{n_{i+k-1}} \left[\sum_{p=1}^{n_i-1} \lg p(x_{ip}) + \sum_{j=1}^k \lg q_j \right] = - \frac{1}{n_{i+k-1}} \left[\sum_{p=1}^{n_i-1} \lg p(x_{ip}) + / \right. \\
& \left. + k \cdot \lg p(x_{in_i}) - k \cdot \lg p(x_{in_i}) + \sum_{j=1}^k \lg q_j \right] = - \frac{1}{n_{i+k-1}} \left[\sum_{p=1}^{n_i-1} \lg p(x_{ip}) + / \right. \\
& \left. + (k-1) \cdot \lg p(x_{in_i}) + \sum_{j=1}^k \lg \left(\frac{q_j}{p(x_{in_i})} \right) \right], \text{ teniendo en cuenta las defini- /} \\
& \text{ciones de } D_{n_i} \text{ y } D_k \text{ se obtiene (en forma abreviada):}
\end{aligned}$$

$$(n_{i+k-1}) \cdot D_{n_{i+k-1}} = n_i D_{n_i} + k \cdot D_k + (k-1) \cdot \lg p(x_{in_i}), \text{ que es e- /}$$

$$\text{quivalente a } D_{n_{i+k-1}} = \frac{n_i}{n_{i+k-1}} \cdot D_{n_i} + \frac{k}{n_{i+k-1}} \cdot D_k + \frac{k-1}{n_{i+k-1}} \cdot \lg p(x_{in_i}).$$

IV_D¹) La propiedad IV_D puede establecerse en una forma equivalen-
te utilizando la notación multiplicativa de Aczel y Daroczy /
(Wang y Suen, (3)) según:

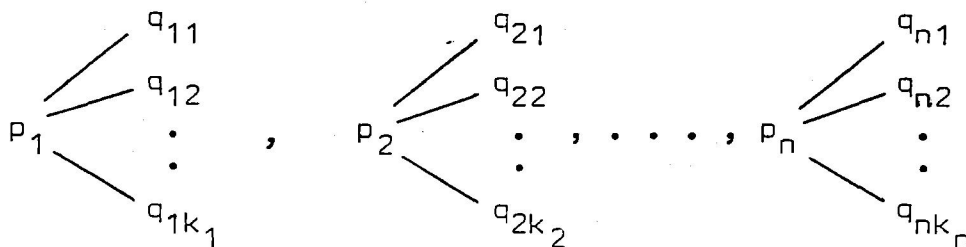
$$\begin{aligned}
& D_{n_{i+k-1}}(p(x_{i1}), \dots, p(x_{in_i}), p(x_{in_i}) \cdot q_1, \dots, p(x_{in_i}) \cdot q_k) = / \\
& = \frac{n_i}{n_{i+k-1}} \cdot D_{n_i}(p(x_{i1}), \dots, p(x_{in_i})) + \frac{k}{n_{i+k-1}} \cdot D_k(q_1, \dots, q_k) + / \\
& + \frac{k-1}{n_{i+k-1}} \cdot \lg p(x_{in_i}), \text{ si } \sum_{j=1}^k q_j = 1.
\end{aligned}$$

para demostrar IV_D¹) basta tomar $Q_j = p(x_{in_i}) \cdot q_j, j=1, \dots, k /$

$$\text{pues } \sum_{j=1}^k Q_j = \sum_{j=1}^k p(x_{in_i}) \cdot q_j = p(x_{in_i}) \cdot \sum_{j=1}^k q_j = p(x_{in_i}) \text{ y aplicando la /}$$

propiedad IV_D a Q_1, \dots, Q_k se obtiene IV'_D .

Las propiedades IV_D) y IV'_D) pueden generalizarse, para cada variable, en caso de que desdoblen varios o todos los estados de dicha variable. Simplificando n_i por n , $p(x_{ij})$ por p_j y D_{n_i} por D el desdoblamiento



origina el teorema:

3.4 Teorema.

Si se verifica $\sum_{j=1}^{k_i} q_{ij} = 1, i=1, \dots, c, 1 \leq c \leq n$, entonces /

$$(n + \sum_{i=1}^c k_i - c) D(p_1 q_{11}, \dots, p_1 q_{1k_1}, p_2 q_{21}, \dots, p_2 q_{2k_2}, \dots, p_c q_{c1}, \dots, p_n) = /$$

$$= n \cdot D(p_1, \dots, p_n) + \sum_{i=1}^c k_i D(q_{i1}, \dots, q_{ik_i}) + \sum_{i=1}^c (k_i - 1) \cdot \lg p_i.$$

por la propiedad II_D) se puede aplicar IV'_D) a cualquier estado, /
lo que permite realizar la demostración por inducción:

- para $c=1$ basta aplicar sucesivamente II_D) y IV'_D).
- para $c=2$ se tiene:

aplicando IV'_D) a $n+k_2-1$ se llega a la expresión /

$$(n+k_1+k_2-2) D(p_1 q_{11}, \dots, p_1 q_{1k_1}, p_2 q_{21}, \dots, p_2 q_{2k_2}, p_3, \dots, p_n) = /$$

$$(n+k_2-1) D(p_1, p_2 q_{21}, \dots, p_2 q_{2k_2}, p_3, \dots, p_n) + k_1 D(q_{11}, \dots, q_{1k_1}) + (k_1-1) \lg p_1.$$

Repitiendo la aplicación de IV'_D), ahora al estado p_2 se llega a /
que la expresión de partida se transforma en:

$$n \cdot D(p_1, \dots, p_n) + k_2 D(q_{21}, \dots, q_{2k_2}) + (k_2 - 1) \cdot \lg p_2 + k_1 D(q_{11}, \dots, q_{1k_1}) + (k_1 - 1) \cdot \lg p_1. \quad /$$

- suponiendo válida la afirmación para c , la validez para $c+1$ se consigue según:

sea $A = D(p_1 q_{11}, \dots, p_1 q_{1k_1}, \dots, p_{c+1} q_{c+1,1}, \dots, p_{c+1} q_{c+1, k_{c+1}})$. Entonces/

$$(n + \sum_{i=1}^{c+1} k_i - (c+1)) \cdot A = (n + k_{c+1} - 1 + \sum_{i=1}^c k_i - c) \cdot A = T = \quad /$$

$$(n + k_{c+1} - 1) D(p_1, \dots, p_c, p_{c+1} q_{c+1,1}, \dots, p_{c+1} q_{c+1, k_{c+1}}, p_{c+2}, \dots, p_n) + \quad /$$

$$+ \sum_{i=1}^c k_i \cdot D(q_{i1}, \dots, q_{ik_i}) + \sum_{i=1}^c (k_i - 1) \cdot \lg p_i, \text{ lo que resulta de aplicar la}$$

hipótesis de inducción. Aplicando IV'_D) al estado p_{c+1} se obtiene /

$$T = n \cdot D(p_1, \dots, p_n) + k_{c+1} D(q_{c+1,1}, \dots, q_{c+1, k_{c+1}}) + (k_{c+1} - 1) \cdot \lg p_{c+1} + \quad /$$

$$+ \sum_{i=1}^c k_i D(q_{i1}, \dots, q_{ik_i}) + \sum_{i=1}^c (k_i - 1) \cdot \lg p_i, \text{ de donde se llega a} \quad /$$

$$T = n \cdot D(p_1, \dots, p_n) + \sum_{i=1}^{c+1} k_i D(q_{i1}, \dots, q_{ik_i}) + \sum_{i=1}^{c+1} (k_i - 1) \cdot \lg p_i.$$

3.5 Corolario.

Si $\sum_{j=1}^{k_i} q_{ij} = 1, i=1, \dots, n$ entonces se verifica /

$$\left(\sum_{i=1}^n k_i \right) D(p_1 q_{11}, \dots, p_1 q_{1k_1}, p_2 q_{21}, \dots, p_n q_{n1}, \dots, p_n q_{nk_n}) =$$

$$= n \cdot D(p_1, \dots, p_n) + \sum_{i=1}^n k_i D(q_{i1}, \dots, q_{ik_i}) + \sum_{i=1}^n (k_i - 1) \cdot \lg p_i.$$

para demostrarlo basta tomar $c=n$ en el teorema ya que entonces/

se obtiene que $(n + \sum_{i=1}^n k_i - n) = \sum_{i=1}^n k_i.$

El teorema y su corolario pueden aplicarse al estudio de árboles/ de clasificación sin solapamiento (Wang y Suen, (3), lo realizan a / partir de la entropía, buscando la disminución de entropía según se / va descendiendo a lo largo del árbol). En este caso el factor /

$$\sum_{i=1}^n k_i \quad \text{ó} \quad \sum_{i=1}^c k_i \quad \text{es el número de clases } c_0$$

respondientes a un nodo interno, n es el número de nodos hijos de es- te nodo, k_i el número de hijos de cada nodo hijo, p_i son las probabi- lidades correspondientes a los nodos hijos y q_i las de los hijos de / los nodos hijos.

La agrupación de varias variables con igual número de estados en/ una variable "media" conlleva una pérdida en la información que pro- / porcionan, es decir, la distorsión producida por la media es menor / que la media de las distorsiones de las variables originales (intuiti- vamente, cuanto mayor sea el número de características que describen/ una distribución más información se obtendrá debido al aumento del de- talle en la descripción). Así, sea $n=n_1=n_2=\dots=n_k$ se puede formar la variable X con los estados x_1, x_2, \dots, x_k en la que la frecuencia de ca- da estado se constituye como la media aritmética de las frecuencias / de los estados correspondientes en las k variables X_1, \dots, X_k , esto es

$$P(X_j) = \frac{p(x_{1j}) + p(x_{2j}) + \dots + p(x_{kj})}{k} \quad \text{ó, en forma simplificada,}$$

$$X = \frac{X_1 + X_2 + \dots + X_k}{k}$$

La relación entre la distorsión de X y las distorsiones de las variables X_1, X_2, \dots, X_k la cuantifica el teorema siguiente:

3.6 Teorema.

Sean las variables X_1, X_2, \dots, X_k tal que $n_1 = \dots = n_k = n$. Entonces /

$$D\left(\frac{X_1 + X_2 + \dots + X_k}{k}\right) \leq \frac{D(X_1) + D(X_2) + \dots + D(X_k)}{k} .$$

del análisis elemental se tiene que $\frac{1}{k} \cdot \sum_{j=1}^k y_j \geq \sqrt[k]{\prod_{j=1}^k y_j}$ que, /

aplicado a las frecuencias de los estados de X se convierte en /

$$p(X_j) = \frac{1}{k} \cdot \sum_{i=1}^k p(x_{ij}) \geq \sqrt[k]{\prod_{i=1}^k p(x_{ij})} , \text{ de donde, por el crecimiento de /}$$

la función logaritmo se obtiene $\lg p(x_j) \geq \lg \sqrt[k]{\prod_{i=1}^k p(x_{ij})} =$ /

$$= \frac{1}{k} \cdot \sum_{j=1}^k \lg p(x_{ij}) . \text{ De aquí, teniendo en cuenta que } X_j = \bar{x}_{ij} \text{ se tiene /}$$

$$D\left(\frac{X_1 + \dots + X_k}{k}\right) = -\frac{1}{n} \cdot \sum_{j=1}^n \lg p(x_j) = -\frac{1}{n} \cdot \sum_{j=1}^n \lg \frac{p(x_{1j}) + \dots + p(x_{kj})}{k} \leq /$$

$$\leq \frac{1}{k} \cdot \left[-\frac{1}{n} \cdot \sum_{j=1}^n \lg p(x_{1j}) - \frac{1}{n} \cdot \sum_{j=1}^n \lg p(x_{2j}) - \dots - \frac{1}{n} \cdot \sum_{j=1}^n \lg p(x_{kj}) \right] = /$$

$$= \frac{D(x_1) + \dots + D(x_k)}{k} .$$

Cuando las frecuencias de los estados correspondientes en cada variable son iguales ($p(x_{1j}) = \dots = p(x_{kj})$, $j=1, \dots, n$), se verifica / la igualdad $D\left(\frac{X_1 + \dots + X_k}{k}\right) = \frac{D(x_1) + \dots + D(x_k)}{k}$, puesto que entonces no / se distinguen las variables en función de las frecuencias de sus esta dos y es posible considerar la variable media que sumariza toda la in formación respecto a dichas variables.

CAPITULO II

Los elementos de una distribución de elementos influyen en la distorsión del espacio aportando, cada uno de ellos, su propia distorsión y modificando por consiguiente la estructura del espacio puesto que la introducción o eliminación de una unidad acarrea una modificación estructural del espacio total en función de las características propias del elemento. La medida de esta influencia debe ser un parámetro acorde con los requisitos I), II) y III) que tiene que verificar la función de evaluación de la entropía ó distorsión, incorporando la información suministrada por los estados que aparecen en la unidad, / su frecuencia y el número de estados de las variables.

1. DEFINICIONES PREVIAS.

1.1 Espacio de unidades.

Sea el conjunto de elementos $\Delta = \{A_1, A_2, \dots, A_n\}$ definido sobre / el conjunto de variables multiestado $V = \{X_1, X_2, \dots, X_v\}$, donde cada X_i tiene asociado un conjunto de estados $W_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$ co- / rrespondiente a los estados que toman las unidades en dicha variable.

A cada unidad A_i se le asocia un objeto matemático $m(A_i)$ definido por la n-tupla (Dubois y Prade, (2), pp. 317-318):

$$m(A_i) = (m_1(A_i), m_2(A_i), \dots, m_v(A_i))$$

siendo m_k el procedimiento de medida asociado a la variable X_k y el / término $m_k(A_i)$ es el valor de la unidad A_i en la variable X_k / $k=1, \dots, v$, $i=1, \dots, n$.

Se define el término $m_k(A_i)$ según:

dado $A_i \in \Delta$, entonces $m_k(A_i) = x_{kj}$ si j es el índice del estado de la variable X_k que aparece en la unidad A_i .

El conjunto de los objetos matemáticos correspondiente a una dis-

tribución de unidades se denominará espacio de patrones (pattern space) S ó espacio total. La diferencia entre una unidad A y su objeto / matemático asociado $m(A)$ es evidente pero, por simplicidad de notación, se denotarán de igual forma (es decir, $S \equiv \Delta$) realizando la distinción cuando sea necesario.

1.2 Probabilidad de una unidad en una variable.

Cada unidad A toma un estado en la variable X_i determinado univocamente por $m_i(A)$ mediante la función

$$m_i: S \longrightarrow W_i, \quad 1 \leq i \leq v.$$

A partir del objeto matemático $m(A)$ (definido por las funciones m_i) / se definen las probabilidades p_i , una para cada variable X_i , en la / forma:

$$p_i: \Delta \longrightarrow [0,1] \text{ siendo } p_i(m(A)) = p(m_i(A)), \quad 1 \leq i \leq v, \quad /$$

es decir, la probabilidad del objeto $m(A)$ (ó de la unidad A) en la va / riante X_i es la probabilidad de $m_i(A)$. En forma simplificada se escribi / rá $p_i(m(A)) = p_i(A)$.

El valor $p_i(A)$ es la frecuencia-probabilidad (frecuencia relativa / ó probabilidad en un diseño probabilístico) del estado x_{ij} si la uni / dad A toma dicho estado en la variable X_i :

$$p_i(A) = p(x_{ij}), \quad 1 \leq j \leq n_i \quad \text{si} \quad m_i(A) = x_{ij}.$$

1.3 Campo-perturbación de una unidad.

La distorsión producida por una unidad origina una perturbación en / el espacio determinada por la información que aporta cada estado que / aparece en la unidad.

El campo asociado a una unidad A tiene que computar la informa- / ción que aporta dicha unidad al espacio total distorsionando la es- / tructura del mismo en función de los estados que lo determinen.

Se define el campo asociado a una unidad A como:

$$F(A) = - \sum_{i=1}^v \frac{1}{n_i} \cdot \lg p_i(A), \quad A \in \Delta,$$

ó en términos de información $F(A) = - \sum_{i=1}^v I(m_i(A)).$

Por construcción, el campo de una variable verifica las propiedades siguientes:

- A) simetría. El campo $F(A)$ es invariante respecto al orden en / que se consideren las variables y los estados de las mismas.
- B) continuidad. La función F es continua en cada uno de los parámetros p_i debido a la continuidad de la función $\lg x$ (la demostración es paralela a la que aparece en Reza, (1), pp. 83).

2. TEOREMAS DEL CAMPO DE UNA UNIDAD.

La introducción ó eliminación de una unidad produce una modificación en la estructura de la distribución, que se comporta como un campo de fuerzas en equilibrio dinámico (p.e. un campo gravitatorio). El campo (ó distorsión) asociado a la unidad produce un reajuste en los valores de los campos de las restantes unidades puesto que influye en las probabilidades de cada uno de los estados y, por tanto, en la relación entre unidades, entre los estados de las variables, etc.

Estas modificaciones estructurales se cuantifican en los teoremas siguientes:

2.1 Teorema I.

Sea $A \in \Delta$. Si la variable X_k modifica el número de sus estados / sin alterar $p_k(A)$, se verifica:

- a) si el número de estados n_k aumenta en h estados, el campo de / A disminuye en el valor $R_h(A) = - \frac{h}{n_k(n_k+h)} \cdot \lg p_k(A).$

b) si el número de estados n_k disminuye en h estados, el campo / de A aumenta en el valor $R_h(A) = - \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A)$.

a) por definición, $F(A) = - \frac{1}{n_1} \lg p_1(A) - \dots - \frac{1}{n_{k-1}} \lg p_{k-1}(A) - \frac{1}{n_k} \lg p_k(A) - \frac{1}{n_{k+1}} \lg p_{k+1}(A) - \dots - \frac{1}{n_v} \lg p_v(A)$. Al aumentar n_k a n_k+h manteniéndose $p_k(A)=cte$. se tiene que la expresión del nuevo campo de A es $F'(A) = - \frac{1}{n_1} \lg p_1(A) - \dots - \frac{1}{n_{k-1}} \lg p_{k-1}(A) - \frac{1}{n_k+h} \lg p_k(A) - \frac{1}{n_{k+1}} \lg p_{k+1}(A) - \dots - \frac{1}{n_v} \lg p_v(A)$. Entonces el incremento en el campo de A es:

$$\Delta F(A) = F(A) - F'(A) = - \left(\frac{1}{n_k} - \frac{1}{n_k+h} \right) \cdot \lg p_k(A) = - \frac{h}{n_k(n_k+h)} \lg p_k(A),$$

y $\Delta F(A) > 0$ puesto que $\lg p_k(A) < 0$ al ser $0 < p_k(A) < 1$. De aquí se deduce que $F'(A) = F(A) - R_h(A)$, siendo $R_h(A) = - \frac{h}{n_k(n_k+h)} \lg p_k(A) > 0$.

b) con un razonamiento paralelo a a) se llega a que, en este caso $\Delta F(A) = F(A) - F'(A) = - \left(\frac{1}{n_k} - \frac{1}{n_k-h} \right) \cdot \lg p_k(A) = \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A)$. Y $\Delta F(A) < 0$ por lo que $F'(A) = F(A) + R'_h(A)$, donde $R'_h(A) = - \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A) > 0$.

El teorema expresa que si una variable experimenta un aumento / (disminución) en el número de sus estados de forma que la frecuencia / de algunos de ellos permanezca constante, el campo de las unidades que posean algunos de dichos estados disminuye (aumenta) en la cantidad R_h (R'_h). Este resultado asevera que se puede disminuir la perturbación originada por una unidad A aumentando el número de estados de / alguna variable X_k (esto es, subdividiendo uno o varios estados, añadiendo nuevos, etc.).

2.2 Corolario I_a.

Sea $A \in \Delta$. Si el número de estados de X_k aumenta a n_k+h quedando $p_k(A)$ constante, el límite del nuevo campo de A es

$$F'(A) = F(A) + \frac{1}{n_k} \cdot \lg p_k(A).$$

del teorema I se tiene que $\Delta F(A) = -\frac{h}{n_k(n_k+h)} \cdot \lg p_k(A)$, que es una sucesión creciente en h pues del análisis elemental se tiene / que si $0 < x < y$ y $a > 0$ entonces $\frac{x}{a(x)} < \frac{y}{a(y)}$ por lo que se llega a

$$-\frac{x}{n_k(n_k+x)} > -\frac{y}{n_k(n_k+y)} \quad \text{y} \quad -\frac{x}{n_k(n_k+x)} \cdot \lg p_k(A) < -\frac{y}{n_k(n_k+y)} \cdot \lg p_k(A).$$

Además $\lim_{h \rightarrow \infty} \frac{h}{n_k(n_k+h)} = \frac{1}{n_k}$ por lo que $\lim_{h \rightarrow \infty} \Delta F(A) = -\frac{1}{n_k} \lg p_k(A)$.

Este valor constituye pues una cota superior para $\Delta F(A)$ y, por tanto el mínimo para $F'(A)$ es $F(A) + \frac{1}{n_k} \cdot \lg p_k(A)$.

El corolario indica que un aumento "ad infinitum" de n_k (en las condiciones del teorema 2.1 no es muy significativo respecto a un aumento más moderado. En particular, para incrementos h_1 y h_2 de n_k , este aumento viene dado por

$$S = F_{h_1}(A) - F_{h_2}(A) = \left(\frac{h_2}{n_k(n_k+h_2)} - \frac{h_1}{n_k(n_k+h_1)} \right) \cdot \lg p_k(A),$$

$h_1 < h_2$. La complicación en la estructura de la distribución es mayor/ que la disminución en $F(A)$ hasta un valor umbral h_0 .

El apartado b) del teorema 2.1 admite un razonamiento análogo aun que restringido a valores de h comprendidos entre 1 y n_k-2 (si fuese/ $h = n_k-1$ la variable X_k quedaría con un sólo estado de probabilidad 1 y no sería informativa ya que todas las unidades poseerían dicho estado).

2.3 Corolario I_b.

Sea $A \in \Delta$. Si el número de estados de X_k disminuye de n_k a n_k-h , / $1 \leq h \leq n_k-2$ permaneciendo $p_k(A)$ constante, entonces $F'(A)$ posee un míni-

mo $m = F(A) - \frac{1}{n_k(n_k-1)} \cdot \lg p_k(A)$, y un máximo $M = F(A) - \frac{n_k-2}{2 \cdot n_k} \lg p_k(A)$.

del teorema 2.1 se obtiene que $\Delta F(A)$ es una sucesión decreciente en h ya que del análisis se tiene que si $a > 0$ y $0 < x < y < a$, entonces $\frac{x}{a(a-x)} < \frac{y}{a(a-y)}$, esto es, $\frac{x}{n_k(n_k-x)} < \frac{y}{n_k(n_k-y)}$. Al ser $p_k(A) < 1$ se deduce que $\frac{x}{n_k(n_k-x)} \cdot \lg p_k(A) > \frac{y}{n_k(n_k-y)} \cdot \lg p_k(A)$, es decir, $\Delta F(A)$ es monótona decreciente. Entonces el máximo de $\Delta F(A)$ será para $h=1$, esto es

$M = F(A) - \frac{n_k-2}{2 \cdot n_k} \cdot \lg p_k(A)$ y el mínimo se alcanza para el valor $h=n_k-2$ en

$$m = F(A) - \frac{1}{n_k(n_k-1)} \cdot \lg p_k(A).$$

La igualdad $m = M$ sólo ocurre en el caso de que $n_k-2 = 1$, es decir si $n_k=3$ pues:

si $n = M$ entonces $\frac{n_k-2}{2 \cdot n_k} = \frac{1}{n_k(n_k-1)}$, de donde $n_k^2 - 3 \cdot n_k = 0$ por

lo que $n_k = 3$ (la solución $n_k = 0$ no tiene sentido).

Para dos decrementos $h_1 > h_2$ el aumento producido en los nuevos está dado por

$$S = F_{h_1}(A) - F_{h_2}(A) = \left(\frac{h_1}{n_k(n_k-h_1)} - \frac{h_2}{n_k(n_k-h_2)} \right) \cdot \lg p_k(A) = \frac{h_1 - h_2}{(n_k-h_1)(n_k-h_2)} \cdot \lg p_k(A)$$

(la disminución en el corolario 2.2 tenía una estructura semejante $\frac{h_2 - h_1}{(n_k+h_2)(n_k+h_1)} \cdot \lg p_k(A)$). Una disminución

drástica de n_k consigue un aumento muy significativo del campo de A pero origina una pérdida de detalle en la descripción de la distribución muy significativa.

2.4 Teorema II

Sea $A \in \Delta$. Si la variable X_k modifica la frecuencia de sus estados sin alterar n_k se verifica:

a) si $p_k(A)$ aumenta en $\Delta p_k(A)$, el campo de A disminuye en la

cantidad $R(A) = \frac{1}{n_k} \cdot \lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right)$.

b) si $p_k(A)$ disminuye en $\Delta p_k(A)$, el campo de A aumenta en la /

cantidad $R'(A) = -\frac{1}{n_k} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right)$.

a) por definición se tiene que $F(A) = -\sum_{i=1}^v \frac{1}{n_i} \cdot \lg p_i(A)$ y, /

como $n_k = \text{cte}$. $F'(A) = -\sum_{i \neq k}^v \frac{1}{n_i} \cdot \lg p_i(A) - \frac{1}{n_k} \cdot \lg(p_k(A) + \Delta p_k(A))$ cons-

tituye la expresión del nuevo campo de A. Entonces $\Delta F(A) = F(A) - /$

$- F'(A) = \frac{1}{n_k} \cdot \lg(p_k(A) + \Delta p_k(A)) - \frac{1}{n_k} \cdot \lg p_k(A) = /$

$= \frac{1}{n_k} \cdot \lg \left(\frac{p_k(A) + \Delta p_k(A)}{p_k(A)} \right) = \frac{1}{n_k} \cdot \lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right) > 0$. Por tanto se lle

ga a $F'(A) = F(A) - R(A)$, siendo $R(A) = \frac{1}{n_k} \cdot \lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right) > 0$, es de

cir, $F'(A) < F(A)$.

b) el nuevo campo de A tiene la expresión determinada por /

$F'(A) = -\sum_{i \neq k}^v \frac{1}{n_i} \cdot \lg p_i(A) - \frac{1}{n_k} \cdot \lg(p_k(A) - \Delta p_k(A))$ por lo que el in

cremento es $F(A) = \frac{1}{n_k} \cdot (\lg(p_k(A) - \Delta p_k(A)) - \lg p_k(A)) = /$

$= \frac{1}{n_k} \cdot \lg \left(\frac{p_k(A) - \Delta p_k(A)}{p_k(A)} \right) = \frac{1}{n_k} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) < 0$. Por tanto se /

deduce $F'(A) = F(A) + R'(A)$, siendo $R'(A) = -\frac{1}{n_k} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) > 0$, /

es decir, $F'(A) > F(A)$.

El aumento (disminución) en el campo de una unidad llega a su extremo cuando las hipótesis del teorema se llevan al límite:

2.5 Corolario II_a.

Sea $A \in \Delta$. Si la variable X_k aumenta su probabilidad $p_k(A)$, sin modificar n_k , en $\Delta p_k(A)$ la disminución $R(A)$ tiene un mínimo en $m = 0$ y un máximo en $M = \frac{1}{n_k} \cdot \lg p_k(A)$.

la función $f(x) = \lg(1 + \frac{x}{p})$ es continua al serlo $\lg x$, y creciente pues si $x_1 < x_2$ se tiene que $f(x_1) - f(x_2) = \lg \left(\frac{p+x_1}{p+x_2} \right)$ y como $p+x_1 < p+x_2$ entonces $\frac{p+x_1}{p+x_2} < 1$ y $\lg \left(\frac{p+x_1}{p+x_2} \right) < 0$, es decir, $f(x_1) < f(x_2)$. Como $0 < \Delta p_k(A) < 1 - p_k(A)$ pues $p_k(A) + \Delta p_k(A) < 1$, el máximo de $R(A)$ se alcanza con $\Delta p_k(A) = 1 - p_k(A)$ y el mínimo tomando $\Delta p_k(A) = 0$, es decir, $M = \frac{1}{n_k} \cdot \lg \left(1 + \frac{1 - p_k(A)}{p_k(A)} \right) = \frac{1}{n_k} \cdot \lg \frac{1}{p_k(A)} = -\frac{1}{n_k} \cdot \lg p_k(A)$, y $m = 0$.

Estos valores máximo y mínimo son intuitivamente evidentes ya que si $\Delta p_k(A) = 0$ se tiene que $R(A) = 0$ y $F'(A) = F(A)$ indicando que la frecuencia de X_k no se ha modificado. En el caso de $\Delta p_k(A) = 1 - p_k(A)$ se verifica que $p_k(A) + \Delta p_k(A) = 1$ por lo que proporcionará una información nula la nueva X_k ya que estará presente en todas las unidades.

2.6 Corolario II_b.

Sea $A \in \Delta$. Si la variable X_k disminuye su probabilidad $p_k(A)$ en $\Delta p_k(A)$ sin modificar n_k , el aumento $R'(A)$ en el campo de A no tiene cota superior.

dato $M > 0$, para que $R'(A) > M$ debe verificarse que $-\frac{1}{n_k} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) > M$, es decir, $\lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) < -M \cdot n_k$ o lo que es equivalente:

$\lg \frac{p_k(A) - \Delta p_k(A)}{p_k(A)} < -M \cdot n_k$. Por el crecimiento de $\lg x$ y tomando an-

tilogaritmos se tiene que $\frac{p_k(A) - \Delta p_k(A)}{p_k(A)} < 2^{-M \cdot n_k}$ esto es /

$p_k(A) - \Delta p_k(A) < 2^{-M \cdot n_k} \cdot p_k(A)$ y $\Delta p_k(A) > p_k(A) - 2^{-M \cdot n_k} \cdot p_k(A)$ /

$= p_k(A) \left(1 - 2^{-M \cdot n_k}\right)$. Por tanto, dado $M > 0$ basta tomar el incremento /

según $\Delta p_k(A) > p_k(A) \cdot \left(1 - 2^{-M \cdot n_k}\right)$ para que se verifique $R'(A) > M$. E-

videntemente $\Delta p_k(A)$ permanece en el rango $0 < \Delta p_k(A) < p_k(A)$ puesto /

que $p_k(A) \cdot \left(1 - 2^{-M \cdot n_k}\right) < p_k(A)$ al ser $2^{-M \cdot n_k} = \frac{1}{2^{M \cdot n_k}} < 1$.

Intuitivamente, al ser $p_k(A) - \Delta p_k(A) \rightarrow 0$, la frecuencia del / estado correspondiente se acerca a 0 produciendo un gran aumento en / la información que proporciona ("ad infinitum").

Las variaciones simultáneas en el número y la frecuencia de los / estados de una variable originan cambios en el campo de las unidades / que se determinan con detalle en el teorema siguiente:

2.7 Teorema III

Sea $A \in \Delta$. Si la variable X_k modifica el número de sus estados y / se altera $p_k(A)$, entonces:

a) si n_k aumenta en h estados y $p_k(A)$ aumenta hasta el valor /

$p_k(A) + \Delta p_k(A) < 1$ el campo de A disminuye a

$$F'(A) = F(A) + \frac{h}{n_k(n_k+h)} \cdot \lg p_k(A) - \frac{1}{n_k+h} \cdot \lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)}\right).$$

b) si n_k aumenta en h estados y $p_k(A)$ disminuye hasta el valor /

$0 < p_k(A) - \Delta p_k(A) < 1$, el campo de A disminuye hasta

$$F'(A) = F(A) + \frac{h}{n_k(n_k+h)} \cdot \lg p_k(A) - \frac{1}{n_k+h} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)}\right).$$

c) si n_k disminuye en h estados y $p_k(A)$ aumenta hasta el valor /
 $p_k(A) + \Delta p_k(A) < 1$, el campo de A se transforma en

$$F'(A) = F(A) - \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A) - \frac{1}{n_k-h} \cdot \lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right).$$

d) si n_k disminuye en h estados y $p_k(A)$ disminuye hasta el valor
 $0 < p_k(A) - \Delta p_k(A)$, el campo de A aumenta hasta

$$F'(A) = F(A) - \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A) - \frac{1}{n_k-h} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right).$$

Demostración

a) por definición $F(A) = - \sum_{i=1}^v \frac{1}{n_i} \cdot \lg p_i(A)$, y el nuevo campo de /

es $F'(A) = - \sum_{i \neq k}^v \frac{1}{n_i} \cdot \lg p_i(A) - \frac{1}{n_k+h} \cdot \lg (p_k(A) + \Delta p_k(A))$. El incremento /

en el campo será $\Delta F(A) = F(A) - F'(A) = \frac{1}{n_k+h} \cdot \lg (p_k(A) + \Delta p_k(A)) - /$
 $-\frac{1}{n_k} \cdot \lg p_k(A)$, y como $\frac{1}{n_k} = \frac{1}{n_k+h} + \frac{h}{n_k(n_k+h)}$ se deduce que el incre-

mento es $\Delta F(A) = \frac{1}{n_k+h} \cdot \lg (p_k(A) + \Delta p_k(A)) - \frac{1}{n_k+h} \cdot \lg p_k(A) - /$

$-\frac{h}{n_k(n_k+h)} \cdot \lg p_k(A) = \frac{1}{n_k+h} \cdot \left(\lg (p_k(A) + \Delta p_k(A)) - \lg p_k(A) \right) - /$

$-\frac{h}{n_k(n_k+h)} \cdot \lg p_k(A) = \frac{1}{n_k+h} \cdot \lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right) - \frac{h}{n_k(n_k+h)} \cdot \lg p_k(A)$, y des-

pejando $F'(A)$ se obtiene el enunciado del teorema.

Trivialmente el incremento de $F(A)$ es < 0 pues $\lg p_k(A) < 0$ y /

$\lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right) > 0$ originando una disminución en $F(A)$.

b) : un razonamiento similar al de a) es $\Delta F(A) = F(A) - F'(A) = /$
 $= \frac{1}{n_k+h} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) - \frac{h}{n_k(n_k+h)} \cdot \lg p_k(A)$, por lo que se obtiene la/

$$\text{expresión } F'(A) = F(A) + \frac{h}{n_k(n_k+h)} \cdot \lg p_k(A) - \frac{1}{n_k+h} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) .$$

Al ser $\lg p_k(A) < 0$ y $\lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) < 0$ la tendencia al aumento o disminución en el campo de A vendrá determinada por la dominancia / de uno u otro factor. Habrá aumento si se verifica la desigualdad /

$$\frac{h}{n_k(n_k+h)} \cdot \lg p_k(A) > \frac{1}{n_k+h} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) \text{ que es equivalente a } /$$

$\frac{h}{n_k} \cdot \lg p_k(A) > \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right)$ y disminución si se verifica la desi- /
gualdad en sentido contrario, dominando el aumento en el número de es-
tados sobre la disminución de la probabilidad. Si se cumple la igual-
dad la disminución de $p_k(A)$ se neutraliza con el aumento de n_k .

c) $\Delta F(A) = F(A) - F'(A) = \frac{1}{n_k-h} \cdot \lg (p_k(A) + \Delta p_k(A)) - \frac{1}{n_k} \lg p_k(A) /$
y como $\frac{1}{n_k} = \frac{1}{n_k-h} + \frac{h}{n_k(n_k-h)}$ se llega a la expresión /

$$F(A) = \frac{1}{n_k-h} \cdot \lg (p_k(A) + \Delta p_k(A)) - \frac{1}{n_k-h} \cdot \lg p_k(A) + \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A) =$$

$$= \frac{1}{n_k-h} \cdot \lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right) + \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A)$$
, obteniéndose la expresión/

del teorema al despejar $F'(A)$.

Como $\lg p_k(A) < 0$ y además $\lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right) > 0$, el aumento o dismi-
nución vendrán determinados por la dominancia de un factor u otro. De
forma análoga a b) el aumento tendrá lugar si se cumple /

- $\frac{h}{n_k} \cdot \lg p_k(A) > \lg \left(1 + \frac{\Delta p_k(A)}{p_k(A)} \right)$, dominando el aumento de $p_k(A)$ sobre /
 la disminución de n_k . La dominación de la disminución de n_k sobre el /
 aumento de $p_k(A)$ ocurrirá si se cumple la desigualdad recíproca. El /
 valor de $F(A)$ no varía con la igualdad, al equilibrarse las tenden- /
 cias de n_k y $p_k(A)$.

$$d) \Delta F(A) = F(A) - F'(A) = \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A) + \frac{1}{n_k-h} \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right),$$

$$\text{de donde } F'(A) = F(A) - \frac{h}{n_k(n_k-h)} \cdot \lg p_k(A) - \frac{1}{n_k-h} \cdot \lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) .$$

Como $\lg p_k(A) < 0$ y $\lg \left(1 - \frac{\Delta p_k(A)}{p_k(A)} \right) < 0$, el resultado final es /

un aumento en $F(A)$.

CAPITULO III

La interacción entre unidades determinará la agrupación de las / mismas en función de distintos parámetros. La atracción y la repul- / sión entre unidades junto con el campo entre ellas decidirán, como me- / didas de información, que unidades pueden agruparse y en que grado.

La medida de la atracción entre dos unidades debe reflejar la in- / formación común a ambas en las distintas variables mientras que la re- / pulsión cuantificará la diferencia entre las mismas. Estos conceptos / exigen algunas definiciones previas que proporcionen datos acerca de / las nociones anteriores.

1. INTERSECCION.

1.1 Intersección de dos unidades.

Sean las unidades $A_i, A_j \in \Lambda$, $i, j=1, \dots, p$. Se define la intersec- / ción entre ellas según:

$$A_i \cap A_j = \{a_{kh} / m_k(A_i) = m_k(A_j) = a_{kh}\}, \quad k=1, \dots, v, \quad h=1, \dots, n_k.$$

Evidentemente, $A_i \cap A_j$ no es una unidad del espacio ya que el núme- / ro de estados n_o de $A_i \cap A_j$ verifica $0 \leq n_o \leq v$, y sólo será $n_o = v$ si se / verifica que $A_i \equiv A_j$. La intersección puede considerarse como una sub- / unidad del espacio constituida por los estados de las variables que / coinciden en ambas unidades.

Simbolizando por ϕ la intersección de dos unidades (ó más) que no / tienen ningún estado en común, es decir,

$$A_i \cap A_j = \phi \quad \text{si y sólo si } \forall k, k=1, \dots, v, \quad m_k(A_i) \neq m_k(A_j), \quad /$$

la intersección verifica las siguientes propiedades:

1.2 Propiedades de la intersección.

1) $\forall j=1, \dots, p$ se tiene que $A_j \cap A_j = A_j$.

$$2) A_i \cap A_j = A_j \cap A_i, \quad \forall i, j=1, \dots, p.$$

$$3) (A_i \cap A_j) \cap A_k = A_i \cap (A_j \cap A_k), \quad \forall i, j, k=1, \dots, p.$$

4) $\forall i, j=1, \dots, p$ se verifican las relaciones

$$A_i \cap A_j \subset A_i$$

$$A_i \cap A_j \subset A_j.$$

La demostración de las propiedades es trivial por definición de \cap de intersección.

El concepto de campo de una unidad puede extenderse al objeto matemático $A_i \cap A_j$ con el propósito de obtener la información a ambas unidades.

1.3 Campo de la intersección de dos unidades.

Sean $A_i, A_j \in \Lambda$. El campo asociado a la intersección $A_i \cap A_j$ se define como la información común a A_i y A_j

$$F(A_i \cap A_j) = - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i), \quad \text{si } m_k(A_i) = m_k(A_j).$$

Esta definición es equivalente a:

$$F(A_i \cap A_j) = - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j), \quad \text{si } m_k(A_j) = m_k(A_i) \text{ puesto /}$$

que si $a_{kh} \in A_i \cap A_j$, $h=1, \dots, n_k$ se tiene que $m_k(A_i) = m_k(A_j) = a_{kh}$.

1.4 Propiedades del campo de la intersección.

$$1) F(\phi) = 0.$$

puesto que si $\phi = A_i \cap A_j$ es porque A_i y A_j no tienen ningún elemento en común.

$$2) F(A_i \cap A_i) = F(A_i).$$

ya que $A_i \cap A_i = A_i$.

$$3) F(A_i \cap A_j) = F(A_j \cap A_i).$$

por la conmutatividad de la intersección.

$$4) 0 \leq F(A_i \cap A_j) \leq F(A_i) \quad \text{y} \quad 0 \leq F(A_i \cap A_j) \leq F(A_j).$$

al ser $\lg p(A_i) \leq 0$ está claro que $F(A_i \cap A_j) \geq 0$. Por la propiedad 4) de la intersección se tiene que $F(A_i \cap A_j) \leq F(A_i)$ y $F(A_i \cap A_j) \leq F(A_j)$ ya que los elementos de $A_i \cap A_j$ pertenecen, a la vez, a A_i y A_j .

$$5) F(A_i \cap (A_j \cap A_k)) = F((A_i \cap A_j) \cap A_k).$$

por la asociatividad de la intersección.

$$6) F(A_i) = F(A_i \cap A_j) + \left[- \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) \right] (m_k(A_i) \neq m_k(A_j)), \quad /$$

$j=1, \dots, p.$

$$\begin{aligned} & \text{por definición, } F(A_i) = - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) = \quad / \\ & = - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) (m_k(A_i) \neq m_k(A_j)) = \quad / \\ & = F(A_i \cap A_j) + \left[- \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) \right] (m_k(A_i) \neq m_k(A_j)). \end{aligned}$$

2. UNION.

De forma semejante a la intersección, un conjunto de unidades debe tener asociado un campo que sumarice la información total aportada por todos los elementos del conjunto (ya sea común a todos los elementos, a sólo algunos de ellos ó la información que los diferencia).

2.1 Unión de dos unidades.

Sean $A_i, A_j \in \Lambda$. Se define la unión de A_i y A_j según:

$$A_i \cup A_j = \{x_{kh} / x_{kh} = m_k(A_i) \text{ ó } x_{kh} = m_k(A_j)\}, \quad k=1, \dots, v, \quad / \\ h=1, \dots, n_k.$$

Es evidente que $A_i \cup A_j$ no es una unidad del espacio (ni siquiera u na subunidad) ya que está definida por un número de estados mayor / que v y, por tanto, puede poseer más de una vez a una variable aunque en distintos estados.

2.2 Propiedades de la unión.

- 1) $A_i \cup A_i = A_i, \quad i=1, \dots, p.$
- 2) $A_i \cup A_j = A_j \cup A_i, \quad i, j=1, \dots, p.$
- 3) $A_i \cup (A_j \cup A_k) = (A_i \cup A_j) \cup A_k, \quad i, j, k=1, \dots, p.$
- 4) $A_i \cup (A_j \cap A_k) = (A_i \cup A_j) \cap (A_i \cup A_k), \quad i, j=1, \dots, p.$
- 5) $A_i \cap (A_j \cup A_k) = (A_i \cap A_j) \cup (A_i \cap A_k), \quad i, j, k=1, \dots, p.$
- 6) $A_i \cup \phi = A_i, \quad i=1, \dots, p.$
- 7) $\forall i, j=1, \dots, p$ se verifica:

$$A_i \cup A_j \supset A_i \quad \text{y} \quad A_i \cup A_j \supset A_j.$$

Por definición de unión, la demostración de las propiedades anteriores es inmediata.

De forma análoga a la intersección, el concepto de campo puede ex tenderse a la unión de unidades de forma que sumalice la información/ aportada por todas las unidades del conjunto.

2.3 Campo conjunto de dos unidades.

Sean $A_i, A_j \in \Delta$. El campo conjunto de A_i y A_j se define como:

$$F(A_i \cup A_j) = - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j) \left(m_k(A_i) \neq m_k(A_j) \right).$$

Por definición de $F(A_i)$, el campo conjunto admite la formulación:

$$F(A_i \cup A_j) = F(A_i) + \left[- \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j) \right] \left(m_k(A_i) \neq m_k(A_j) \right).$$

Una definición equivalente se obtiene intercambiando A_i por A_j :

$$F(A_i \cup A_j) = F(A_j) + \left[- \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) \right] \left(m_k(A_i) \neq m_k(A_j) \right).$$

2.4 Propiedades del campo conjunto de dos unidades.

$$1) \quad F(A_i \cup A_j) \geq F(A_i) \quad \text{y} \quad F(A_i \cup A_j) \geq F(A_j).$$

por definición de $F(A_i \cup A_j)$.

$$2) \quad F(A_i \cup A_j) \geq 0.$$

por 1) se tiene que $F(A_i \cup A_j) \geq F(A_i) \geq 0$.

$$3) \quad F(A_i \cap A_j) \leq F(A_i \cup A_j).$$

por la propiedad 6) del campo de la intersección se tiene

$$\begin{aligned} F(A_i \cup A_j) &= F(A_i) + \left[- \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j) \right] \left(m_k(A_i) \neq m_k(A_j) \right) = / \\ &= F(A_i \cap A_j) - \underbrace{\sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j)}_{m_k(A_i) \neq m_k(A_j)} \geq F(A_i \cap A_j). \end{aligned}$$

$$4) F(A_i \cup A_j) \leq F(A_i) + F(A_j).$$

$$\text{como } - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j) \geq - \underbrace{\sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j)}_{m_k(A_i) \neq m_k(A_j)}, \text{ por lo que/}$$

$$F(A_j) \geq - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j) \quad (m_k(A_i) \neq m_k(A_j)) \text{ y al sustituir es-}$$

ta desigualdad en la expresión de $F(A_i \cup A_j)$ se obtiene que /
 $F(A_i \cup A_j) \leq F(A_i) + F(A_j).$

$$5) F(A_i \cup A_j) = F(A_j \cup A_i).$$

por la conmutatividad de la unión.

$$6) F(A_i \cup A_i) = F(A_i).$$

puesto que $A_i \cup A_i = A_i.$

$$7) F(A_i \cup (A_j \cup A_k)) = F((A_i \cup A_j) \cup A_k).$$

por la asociatividad de la unión.

La propiedad 3) tiene una demostración más corta pues:

por la propiedad 4) del campo-intersección y 1) del campo conjun-
to se tiene $F(A_i \cap A_j) \leq F(A_i)$ y $F(A_i) \leq F(A_i \cup A_j).$ Sin embargo, la demos-
tración realizada induce la noción de disimilaridad entre dos unida-
des como medida de la información que se obtiene de las diferencias e
xistentes entre ambas unidades, en base a los estados distintos que /
poseen.

2.5 Disimilaridad entre dos unidades.

Sean $A_i, A_j \in \Lambda.$ La disimilaridad entre A_i y A_j se define como:

$$D(A_i \cup A_j) = - \left[\sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) + \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_j) \right] (m_k(A_i) \neq m_k(A_j)).$$

2.6 Propiedades de la disimilaridad.

1) $D(A_i \cup A_i) = 0.$

puesto que $\forall k=1, \dots, n_k$ se tiene que $m_k(A_i) = m_k(A_i)$ en cada una de las variables.

2) $D(A_i \cup A_j) = F(A_i \cup A_j) - F(A_i \cap A_j).$

trivial por la demostración de la propiedad 3) del campo intersección.

3) $D(A_i \cup A_j) \geq 0.$

al ser $F(A_i \cup A_j) \geq F(A_i \cap A_j)$, por la propiedad 2) anterior se deduce la desigualdad.

4) $D(A_i \cup A_j) \leq F(A_i) + F(A_j).$

por 2) se obtiene que $D(A_i \cup A_j) \leq F(A_i \cup A_j)$, y como la propiedad 4) del campo conjunto afirma que $F(A_i \cup A_j) \leq F(A_i) + F(A_j)$, es inmediata la demostración.

El campo conjunto de dos unidades verifica la propiedad fundamental $F(A) + F(B) = F(A \cup B) + F(A \cap B)$ enunciada por Pal y Majumder ((4), / pp. 62) en el contexto de medir la ambigüedad de un conjunto (estos / autores estudian el grado de fuzziness de un conjunto y formulan la / condición en la forma $d(A) + d(B) = d(A \cup B) + d(A \cap B)$ siendo d la distancia entre los conjuntos A y B). La verificación de la propiedad la asegura el teorema siguiente:

2.7 Teorema A.

Sean $A_i, A_j \in \Delta$. Entonces $F(A_i \cup A_j) = F(A_i) + F(A_j) - F(A_i \cap A_j).$

por definición del campo conjunto de A_i y A_j se tiene que /

$$F(A_i \cup A_j) = F(A_i) + \left[- \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(A_i) \right] (m_k(A_i) \neq m_k(A_j)) = /$$

= $F(A_i) + F(A_j) - F(A_i \cap A_j)$, por la propiedad 6) del campo intersec- /
ción.

2.8 Corolario I_A.

Si $A_i \cap A_j = \emptyset$ entonces $D(A_i \cup A_j) = F(A_i) + F(A_j)$.

como $F(\emptyset) = 0$, se tiene que $F(A_i \cap A_j) = 0$ y, por la propiedad 2) /
de la disimilaridad $D(A_i \cup A_j) = F(A_i \cup A_j)$. Aplicando el Teorema A se ob- /
tiene $D(A_i \cup A_j) = F(A_i) + F(A_j)$.

2.9 Corolario II_A.

$$D(A_i \cup A_j) = F(A_i) + F(A_j) - 2 \cdot F(A_i \cap A_j).$$

aplicando el resultado del teorema A a la propiedad 2) de la /
disimilaridad se tiene que $D(A_i \cup A_j) = F(A_i \cup A_j) - F(A_i \cap A_j) = F(A_i) + /$
 $+ F(A_j) - F(A_i \cap A_j) - F(A_i \cap A_j) = F(A_i) + F(A_j) - 2 \cdot F(A_i \cap A_j)$.

Este teorema permite expresar el campo conjunto de dos unidades /
en función de los campos individuales de las unidades y del campo de /
la intersección, resultado que puede generalizarse al cálculo del cam- /
po conjunto de un grupo de unidades y se obtiene en función de los /
campos individuales de los componentes del grupo y de las sucesivas /
intersecciones entre ellos (dos a dos, tres a tres, etc.) mediante el /
teorema siguiente:

2.10 Teorema B.

Sean $A_1, A_2, \dots, A_n \in \Delta$ las unidades del grupo G_n . Entonces:

$$+ F(A_{n+1}) - \left[\sum_{i=1}^n F(A_i \cup A_{n+1}) - \sum_{i < j} F(A_i \cup A_j \cup A_{n+1}) \right] +$$

ción aplicada a $\bigcup_{i=1}^n (A_i \cup A_{n+1})$ se tiene que $F\left(\bigcup_{i=1}^{n+1} A_i\right) = F\left(\bigcup_{i=1}^n A_i\right) +$

$$= F\left(\bigcup_{i=1}^n A_i\right) + F(A_{n+1}) - F\left(\bigcup_{i=1}^n (A_i \cup A_{n+1})\right); \text{ y, por la hipótesis de induc-}$$

$$F\left(\bigcup_{i=1}^{n+1} A_i\right) = F\left(\bigcup_{i=1}^n A_i\right) \cup A_{n+1} = F\left(\bigcup_{i=1}^n A_i\right) + F(A_{n+1}) - F\left(\bigcup_{i=1}^n (A_i \cup A_{n+1})\right) =$$

- suponiendo la validez para n , para $n+1$ será:

$$- F(A_2 \cup A_3) + F(A_1 \cup A_2 \cup A_3).$$

$$- F(A_1 \cup A_2 \cup A_3) + F(A_1) + F(A_2) + F(A_3) - F(A_1 \cup A_2) - F(A_1 \cup A_3) -$$

$$- F(A_2 \cup A_3) + F(A_1 \cup A_2) + F(A_1 \cup A_3) + F(A_2 \cup A_3) - [F(A_1 \cup A_2) + F(A_1 \cup A_3) +$$

$$F(A_2 \cup A_3)] = F(A_1 \cup A_2 \cup A_3) = F(A_1 \cup A_2) + F(A_3) - F(A_1 \cup A_2) + F(A_3) -$$

de unión e intersección:

- para $n=3$ se tiene que, aplicando el Teorema B y la distributividad

- para $n=2$, el teorema B asegura la validez del enunciado.

ción:

La demostración se realizará mediante el uso del método de induc-

$$\sum_{i=1}^{n-1} F(A_i \cup A_j \cup \dots \cup A_h) \pm F(A_1 \cup A_2 \cup \dots \cup A_n).$$

$$F(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n F(A_i) - \sum_{i < j} F(A_i \cup A_j) + \sum_{i < j < k} F(A_i \cup A_j \cup A_k) - \dots \pm$$

$$+ \sum_{i < j < k}^n F((A_i \cap A_{n+1}) \cap (A_j \cap A_{n+1}) \cap (A_k \cap A_{n+1})) - \dots \pm F\left(\bigcap_1^n (A_i \cap A_{n+1})\right) \Bigg] .$$

Por las propiedades 1), 2) y 3) de la intersección se cumple:

- $(A_i \cap A_{n+1}) \cap (A_j \cap A_{n+1}) = A_i \cap A_j \cap A_{n+1} .$
- $(A_i \cap A_{n+1}) \cap (A_j \cap A_{n+1}) \cap (A_k \cap A_{n+1}) = A_i \cap A_j \cap A_k \cap A_{n+1} .$
-
- $\bigcap_1^n (A_i \cap A_{n+1}) = \bigcap_1^{n+1} A_i .$

Entonces se verifica:

$$F\left(\bigcup_1^{n+1} A_i\right) = F\left(\bigcup_1^n A_i\right) + F(A_{n+1}) - \left[\sum_1^n F(A_i \cap A_{n+1}) - \sum_{i < j}^n F(A_i \cap A_j \cap A_{n+1}) + \right. /$$

$$\left. + \sum_{i < j < k}^n F(A_i \cap A_j \cap A_k \cap A_{n+1}) - \dots \pm F\left(\bigcap_1^{n+1} A_i\right) \right] . \text{ Sustituyendo el valor de } /$$

$F\left(\bigcup_1^n A_i\right)$ en $F\left(\bigcup_1^{n+1} A_i\right)$ de acuerdo a la hipótesis de inducción se obtiene el resultado (agrupando los factores homónimos):

$$F\left(\bigcup_1^{n+1} A_i\right) = \sum_1^n F(A_i) + F(A_{n+1}) - \left[\left(\sum_{i < j}^n F(A_i \cap A_j) + \sum_1^n F(A_i \cap A_{n+1}) \right) - \right. /$$

$$\left. - \left(\sum_{i < j < k}^n F(A_i \cap A_j \cap A_k) + \sum_{i < j}^n F(A_i \cap A_j \cap A_{n+1}) \right) + \dots \pm \left(F\left(\bigcap_1^n A_i\right) + \right. /$$

$$\left. + \sum_{i < j < \dots < h}^n F(A_i \cap A_j \cap \dots \cap A_h \cap A_{n+1}) \right) \pm F\left(\bigcap_1^{n+1} A_i\right) \Bigg] = \sum_1^{n+1} F(A_i) - \sum_{i < j}^{n-1} F(A_i \cap A_j) + /$$

$$+ \sum_{i < j < k}^{n+1} F(A_i \cap A_j \cap A_k) - \dots \pm \sum_{i < j < \dots < h}^{n+1} F(A_i \cap A_j \cap \dots \cap A_h) \pm F\left(\bigcap_1^{n+1} A_i\right).$$

n

2.11 Propiedades de unión e intersección.

$$1) F\left(\bigcap_1^n A_i\right) \leq F(A_j), \quad j=1, \dots, n$$

por la asociatividad y la propiedad 4) del campo de la intersección.

$$F\left(\bigcap_1^n A_i\right) = F\left(\left(\bigcap_{i \neq j}^n A_i\right) \cap A_j\right) \leq F(A_j).$$

$$2) F\left(\bigcap_1^n A_i\right) \leq F\left(\bigcap_{i \neq j}^n A_i\right), \quad j=1, \dots, n.$$

el mismo razonamiento realizado en 1), es decir, aplicación de asociatividad y propiedad 4) del campo de la intersección) / conduce a:

$$F\left(\bigcap_1^n A_i\right) = F\left(\left(\bigcap_{i \neq j}^n A_i\right) \cap A_j\right) \leq F\left(\bigcap_{i \neq j}^n A_i\right).$$

$$3) F\left(\bigcup_1^n A_i\right) \geq 0.$$

por la asociatividad y la propiedad 1) del campo conjunto de dos unidades

$$F\left(\bigcup_1^n A_i\right) = F\left(\left(\bigcup_{i \neq j}^n A_i\right) \cap A_j\right) \geq F(A_j) \geq 0.$$

$$4) F\left(\bigcup_1^n A_i\right) \geq F(A_j), \quad j=1, \dots, n.$$

trivial por el razonamiento de 3).

$$5) F\left(\bigcup_1^n A_i\right) \geq F\left(\bigcup_{i \neq j}^n A_i\right), \quad j=1, \dots, n.$$

el razonamiento de 3) conduce a

$$F\left(\bigcup_1^n A_i\right) = F\left(\left(\bigcup_{i \neq j}^n A_i\right) \cap A_j\right) \geq F\left(\bigcup_{i \neq j}^n A_i\right).$$

$$6) F\left(\bigcap_1^n A_i\right) \leq F\left(\bigcup_1^n A_i\right).$$

combinando las propiedades 1) y 4) se obtiene

$$F\left(\bigcap_1^n A_i\right) \leq F(A_j) \leq F\left(\bigcup_1^n A_i\right), \quad j=1, \dots, n.$$

$$7) F\left(\bigcup_1^n A_i\right) \geq \sum_1^n F(A_i).$$

realizando un razonamiento por inducción, para $n=2$ es trivial por el Teorema A.

Para $n=3$ se tiene que:

$F(A_1 \cup A_2 \cup A_3) = F((A_1 \cup A_2) \cup A_3) = F(A_1 \cup A_2) + F(A_3) - F((A_1 \cup A_2) \cap A_3)$ al aplicar el Teorema A, y por la hipótesis de inducción para $n=2$ se tiene que $F(A_1 \cup A_2) \leq F(A_1) + F(A_2)$, de donde

$$F(A_1 \cup A_2 \cup A_3) \leq \sum_1^3 F(A_i) - F((A_1 \cup A_2) \cap A_3) \leq \sum_1^3 F(A_i), \text{ puesto que se verifica que } F((A_1 \cup A_2) \cap A_3) \geq 0.$$

fica que $F((A_1 \cup A_2) \cap A_3) \geq 0$.

Por la hipótesis de inducción $F\left(\bigcup_1^n A_i\right) \leq \sum_1^n F(A_i)$. Entonces, para

$n+1$ será:

$$F\left(\bigcup_1^{n+1} A_i\right) = F\left(\left(\bigcup_1^n A_i\right) \cup A_{n+1}\right) = F\left(\bigcup_1^n A_i\right) + F(A_{n+1}) - F\left(\left(\bigcup_1^n A_i\right) \cap A_{n+1}\right) \leq /$$

$$\leq \sum_1^{n+1} F(A_i) - F\left(\left(\bigcup_1^n A_i\right) \cap A_{n+1}\right) \leq \sum_1^{n+1} F(A_i).$$

Estas propiedades son la expresión de las ideas intuitivas que reflejan el comportamiento de las unidades respecto a la distorsión que crean en el espacio. En particular, muestran el decrecimiento del campo de la intersección (propiedad 2) al aumentar el número de elementos; la propiedad 5) denota, en iguales circunstancias, el crecimiento del campo conjunto. El hecho de que un conjunto de unidades produce más distorsión en el espacio que una única unidad (propiedad 4)) / induce la definición de la distorsión ó campo que produce un grupo en función de la información que proporcionan sus elementos.

2.12 Campo asociado a un grupo de unidades.

Sea el grupo $G_n = \{A_1, A_2, \dots, A_n\}$. Se define el campo asociado a / según

$$F(G_n) = F\left(\bigcup_1^n A_i\right).$$

Es una medida que computa la información conjunta proporcionada / por los elementos del grupo G_n respecto a la similitud de las unidades que lo componen en las distintas variables así como respecto a / las diferencias existentes entre los estados que las definen, esto es cuantifica tanto la atracción como la repulsión existente entre las n unidades del grupo.

Por definición, el campo asociado a G_n verifica todas las propiedades 1-7 del campo conjunto de varias unidades, entre las que merece destacar la importancia de 5) y 7):

- la propiedad 7) enuncia que el campo de G_n es menor que la suma/ de los campos individuales de sus elementos, de acuerdo a la i-/ dea intuitiva de que la cohesión del grupo hace disminuir la in- certidumbre total de las unidades que lo componen por separado, / en función del grado de asociación que exista entre los elemen- / tos del grupo. Si la asociación entre las unidades es nula (esto es, no poseen ningún estado común) las intersecciones correspon- dientes al Teorema B son todas ϕ y, por consiguiente, se verifi- ca la igualdad

$$F(G_n) = \sum_1^n F(A_i) \text{ lo que significa que la in-}$$

certidumbre del grupo es la suma de las incertidumbres individua- les de sus elementos.

- la propiedad 5) confirma el aumento del campo de G_n al incluir / un elemento nuevo puesto que, en general, modifica la estructura del grupo produciendo un aumento en la incertidumbre (disminuye/ la cohesión) de acuerdo a la disimilaridad que proporcione. Ex- / cepto en el caso de que sus estados coincidan con los estados de elementos de G_n , la nueva unidad disminuye la homogeneidad.

Las propiedades 1), 2) y 6) se refieren a las intersecciones fuer- tes en G_n (la parte común a todas y cada una de las unidades del gru- po), esto es, no contabiliza las similitudes débiles (dos elementos / que coinciden en un sólo estado, etc.) dentro del grupo. Estas asocia- ciones parciales tienen cierto interés y ya aparecen en el Teorema B. En el campo de las medidas de similaridad basadas en Teoría de Conjun- tos Ito, Kodama y Toyoda ((5), pp. 111-112) realizan una generaliza- / ción para encontrar un índice que tenga en cuenta la correlación en- / tre variables no-independientes (aunque solamente correlación dos a / dos entre variables).

CAPITULO IV

La noción de campo de un grupo contiene información tanto de la afinidad (similitud) como de la heterogeneidad (disimilitud) entre sus elementos, reflejando la estructura subyacente en base a estructura / de los estados que los configuran. Estas propiedades estructurales / del grupo pueden expresarse mediante los valores de la afinidad entre un elemento y el grupo ó entre dos grupos (Pal y Majumder (4), pp. / 87). Entre dos elementos la afinidad tiene una caracterización intuitiva: el campo intersección de ambos $A_t(A_1, A_2) = F(A_1 \cap A_2)$, por lo / que se verifica

$$F(A_1 \cup A_2) = F(A_1) + F(A_2) - A_t(A_1, A_2) \text{ equivalente a la expresión } /$$

$$A_t(A_1, A_2) = F(A_1) + F(A_2) - F(A_1 \cup A_2). \text{ Sin embargo, al considerar la } \underline{a}$$

finidad unidad-grupo y grupo-grupo hay que considerar, respectivamente, la atracción que se ejerce entre dicha unidad y los elementos del grupo y la atracción ejercida mutuamente por los elementos de los dos grupos.

1. AFINIDAD.

1.1 Afinidad unidad-grupo.

Sean $A \in \Lambda$ y un grupo G . La afinidad entre A y G se define como

$$A_t(A, G) = F(G \cap A).$$

Enumerando los elementos de G según $G = \{A_1, A_2, \dots, A_n\}$ se tiene / que la afinidad entre A y G toma la forma

$$A_t(A, G) = F\left(\left(\bigcup_1^n A_i\right) \cap A\right).$$

Esta expresión de A_t permite formular la afinidad elemento-grupo / en función de las intersecciones múltiples entre el elemento y los e- / lementos del grupo.

1.2 Teorema C.

Sean $A \in \Lambda$ y el grupo $G = \{A_1, A_2, \dots, A_n\}$. Entonces:

$$A_t(A, G) = \sum_1^n F(A_i \cap A) - \sum_{i < j}^n F(A_i \cap A_j \cap A) + \sum_{i < j < k}^n F(A_i \cap A_j \cap A_k \cap A) - \dots \pm /$$

$$\pm F\left(\left(\bigcap_1^n A_i\right) \cap A\right).$$

para $n=1$ el teorema es trivial.

para $n=2$, aplicando la distributividad se deduce:

$$F((A_1 \cup A_2) \cap A) = F((A_1 \cap A) \cup (A_2 \cap A)) = F(A_1 \cap A) + F(A_2 \cap A) - /$$

$$- F((A_1 \cap A) \cap (A_2 \cap A)) = F(A_1 \cap A) + F(A_2 \cap A) - F(A_1 \cap A_2 \cap A).$$

suponiendo el teorema válido para $n-1$ y aplicando de nuevo la distributividad se tiene para n :

$$A_t(A, G) = F\left(\left(\bigcup_1^n A_i\right) \cap A\right) = F\left[\left(\left(\bigcup_1^{n-1} A_i\right) \cup A_n\right) \cap A\right] = F\left[\left(\left(\bigcup_1^{n-1} A_i\right) \cap A\right) \cup (A_n \cap A)\right] =$$

$$= F\left(\left(\bigcup_1^{n-1} A_i\right) \cap A\right) + F(A_n \cap A) - F\left(\left(\bigcup_1^{n-1} A_i\right) \cap (A_n \cap A)\right).$$

Por la hipótesis de inducción aplicada en A y $A_n \cap A$ se verifican:

$$1) F\left(\left(\bigcup_1^{n-1} A_i\right) \cap A\right) = \sum_1^{n-1} F(A_i \cap A) - \sum_{i < j}^{n-1} F(A_i \cap A_j \cap A) + \dots \pm F\left(\left(\bigcap_1^{n-1} A_i\right) \cap A\right).$$

$$2) -F\left(\left(\bigcup_1^{n-1} A_i\right) \cap (A_n \cap A)\right) = -\sum_1^{n-1} F(A_i \cap (A_n \cap A)) - \sum_{i < j}^{n-1} F(A_i \cap A_j \cap (A_n \cap A)) + /$$

$$+ \dots \pm F\left(\left(\bigcap_1^{n-1} A_i\right) \cap (A_n \cap A)\right).$$

Teniendo en cuenta que en 1) y 2) hay igual número de términos y/

que los signos dentro del corchete de 2) se corresponden con los signos de 1), la sustitución de 1) y 2) en $A_t(A, G)$ permite agrupar los términos en la forma:

$$* \sum_1^{n-1} F(A_i \cap A) + F(A_n \cap A) = \sum_1^n F(A_i \cap A).$$

$$* - \sum_{i < j}^{n-1} F(A_i \cap A_j \cap A) - \sum_1^{n-1} F(A_i \cap A_n \cap A) = - \sum_{i < j}^n F(A_i \cap A_j \cap A).$$

.....

$$* \pm F\left(\left(\bigcap_1^{n-1} A_i\right) \cap A\right) \pm \sum_{\substack{i < j < \dots < k \\ n-2}}^{n-1} F\left(\left(A_i \cap A_j \cap \dots \cap A_k\right) \cap (A_n \cap A)\right) =$$

$$= \pm \sum_{\substack{i < j < \dots < k \\ n-1}}^n F\left(\left(A_i \cap A_j \cap \dots \cap A_k\right) \cap A\right), \text{ (los signos } \pm \text{ coinciden).}$$

$$* \pm F\left(\left(\bigcap_1^{n-1} A_i\right) \cap (A_n \cap A)\right) = \pm F\left(\left(\bigcap_1^n A_i\right) \cap A\right).$$

De aquí se obtiene que $A_t(A, G) = \sum_1^n F(A_i \cap A) - \sum_{i < j}^n F(A_i \cap A_j \cap A) +$ /

$$+ \sum_{i < j < k}^n F(A_i \cap A_j \cap A_k \cap A) - \dots \pm F\left(\left(\bigcap_1^n A_i\right) \cap A\right).$$

De la definición de A_t y el Teorema C surgen las siguientes propiedades relativas a la afinidad:

1.3 Propiedades de la afinidad elemento-grupo.

1) Si $\forall A_i \in G$ se tiene que $A_i \cap A = \emptyset$, entonces $A_t(A, G) = 0$.

a partir de la hipótesis se deduce que las intersecciones de A con cualquier número de elementos de G es \emptyset , es decir, /

$$\forall i, j, \dots, k, \quad A_i \cap A_j \cap \dots \cap A_k \cap A = \emptyset.$$

2) Sean $A \in \Delta$ y el grupo G. Entonces se verifica que:

$$0 \leq A_t(A, G) \leq F(A) \quad \text{y} \quad 0 \leq A_t(A, G) \leq F(G).$$

por definición, $A_t(A, G) = F(G \cap A)$, y por la propiedad 4) / de la intersección se verifica que $F(G \cap A) \leq F(A)$ y / $F(G \cap A) \leq F(G)$.

3) Sean $A \in \Delta$ y el grupo G. Entonces $0 \leq A_t(A, G) \leq F(G \cup A)$.

trivial por la anterior 2) y la propiedad 4) de la unión.

4) Sean $A \in \Delta$ y el grupo G. Entonces $A_t(A, G) = F(G) + F(A) - F(G \cup A)$

sea $G = \{A_1, A_2, \dots, A_n\}$. Por el Teorema B se tiene:

$$\begin{aligned} * F(G \cup A) &= F(A_1 \cup A_2 \cup \dots \cup A_n \cup A) = \sum_1^n F(A_i) + F(A) - \left[\sum_{i < j}^n F(A_i \cap A_j) + \right. / \\ &+ \left. \sum_1^n F(A_i \cap A) \right] + \left[\sum_{i < j < k}^n F(A_i \cap A_j \cap A_k) + \sum_{i < j}^n F(A_i \cap A_j \cap A) \right] - \dots \pm / \\ &\pm \left[F(A_1 \cap A_2 \cap \dots \cap A_n) + \sum_{\substack{i < j < \dots < h \\ n-1}}^n F(A_i \cap A_j \cap \dots \cap A_h \cap A) \right] \pm F\left(\left(\bigcap_1^n A_i\right) \cap A\right). \end{aligned}$$

$$* F(G) = F(A_1 \cap A_2 \cap \dots \cap A_n) = \sum_1^n F(A_i) - \sum_{i < j}^n F(A_i \cap A_j) + /$$

$$+ \sum_{i < j < k}^n F(A_i \cap A_j \cap A_k) - \dots \pm F\left(\bigcap_1^n A_i\right).$$

$$\text{Entonces } F(G) + F(A) - F(G \cup A) = \sum_1^n F(A_i \cap A) - \sum_{i < j}^n F(A_i \cap A_j \cap A) + \dots \pm /$$

$$\pm \sum_{\underbrace{i < j < \dots < h}_{n-1}}^n F(A_i \cap A_j \cap \dots \cap A_h \cap A) \pm F\left(\left(\bigcap_1^n A_i\right) \cap A\right) = A_t(A, G).$$

Esta propiedad se puede expresar según $F(G \cup A) = F(G) + F(A) - A_t(A, G)$ permitiendo calcular el campo del grupo formado a partir del grupo G y el elemento A en función de los campos de A y G y la atracción entre A y G (y viceversa).

La relación existente entre la atracción de un grupo y un elemento A y la atracción del grupo excepto un elemento y A viene determinada por el siguiente teorema.

1.4 Teorema D.

Sean $G_n = \{A_1, A_2, \dots, A_n\}$, $G_{n+1} = \{A_1, A_2, \dots, A_n, A_{n+1}\}$ y $A \in \Delta$. Entonces se verifica:

$$A_t(A, G_{n+1}) = A_t(A, G_n) + F(A_{n+1} \cap A) - \sum_1^n F(A_i \cap A_{n+1} \cap A) + /$$

$$+ \sum_{i < j}^n F(A_i \cap A_j \cap A_{n+1} \cap A) - \dots \pm F\left(\left(\bigcap_1^{n+1} A_i\right) \cap A\right).$$

por el Teorema C se obtienen las expresiones:

$$* A_t(A, G_{n+1}) = \sum_1^{n+1} F(A_i \cap A) - \sum_{i < j}^{n+1} F(A_i \cap A_j \cap A) + \dots \pm F\left(\left(\bigcap_1^{n+1} A_i\right) \cap A\right).$$

$$* A_t(A, G_n) = \sum_1^n F(A_i \cap A) - \sum_{i < j}^n F(A_i \cap A_j \cap A) + \dots - F\left(\left(\bigcap_1^n A_i\right) \cap A\right).$$

Restando las anteriores igualdades término a término se agrupan / los términos en la forma:

$$* \sum_1^{n+1} F(A_i \cap A) - \sum_1^n F(A_i \cap A) = F(A_{n+1} \cap A).$$

$$* - \sum_{i < j}^{n+1} F(A_i \cap A_j \cap A) + \sum_{i < j}^n F(A_i \cap A_j \cap A) = - \sum_1^n F(A_i \cap A_{n+1} \cap A).$$

Los términos n-simos de $A_t(A, G_{n+1})$ y $A_t(A, G_n)$ son, respectiva- / mente

$$\pm \sum_{i < j < \dots < h}^{n+1} F(A_i \cap A_j \cap \dots \cap A_h \cap A) \quad \text{y} \quad \pm F\left(\left(\bigcap_1^n A_i\right) \cap A\right), \text{ ambos con igual sig}$$

no. Al restarlos se obtiene:

$$\pm \sum_{i < j < \dots < h}^n F(A_i \cap A_j \cap \dots \cap A_h \cap A_{n+1} \cap A), \text{ y entonces se obtiene} \quad /$$

n-1

$$A_t(A, G_{n+1}) - A_t(A, G_n) = F(A_{n+1} \cap A) - \sum_1^n F(A_i \cap A_{n+1} \cap A) + \dots \pm F\left(\left(\bigcap_1^{n+1} A_i\right) \cap A\right).$$

A partir del teorema se obtiene el valor máximo de la atracción / (afinidad) elemento-grupo, que ocurre cuando $A \in G$ ya que en este caso A coincide con un elemento de G : él mismo.

1.5 Corolario D1.

Sean el grupo G y $A \in G$. Entonces $A_t(A, G) = F(A)$ y es un máximo /

para la afinidad entre A y cualquier grupo.

sean $G_n = \{A_1, A_2, \dots, A_n\}$ y $G_{n+1} = \{A_1, A_2, \dots, A_n, A_{n+1}\}$. Si el elemento A verifica que $A \equiv A_{n+1}$ se tiene que $A \in G_{n+1}$ y por el Teorema D:

$$A_t(A, G_{n+1}) - A_t(A, G_n) = F(A) - \sum_1^n F(A_i \cap A) + \sum_{i < j}^n F(A_i \cap A_j \cap A) - \dots \pm$$

$$\pm \sum_{i < \dots < h}^n F(\underbrace{A_i \cap \dots \cap A_h}_{n-1} \cap A) \pm F\left(\left(\bigcap_1^n A_i\right) \cap A\right)$$

y, por el Teorema C, la segunda

parte del segundo miembro de la igualdad es $A_t(A, G_n)$, de donde se obtiene $A_t(A, G_{n+1}) - A_t(A, G_n) = F(A) - A_t(A, G_n)$, es decir,

$A_t(A, G_{n+1}) = F(A)$. Y por la propiedad 2) de la afinidad, dado un grupo G' se tiene que $0 \leq A_t(A, G') \leq F(A) = A_t(A, G_{n+1})$, por lo que la afinidad $A_t(A, G_{n+1})$ es un máximo absoluto para $A_t(A, G')$ si $A \in G_{n+1}$.

No es cierta la afirmación recíproca ya que puede ocurrir la igualdad $A_t(A, G) = F(A)$ y ser $A \notin G$ (basta que, entre las coincidencias de A con todos los elementos de G, aparezcan la totalidad de los estados que configuran al elemento A).

1.6 Corolario D2.

Sean los grupos $G_n = \{A_1, \dots, A_n\}$, $G_{n+1} = \{A_1, \dots, A_n, A_{n+1}\}$ y $G'_{n+1} = \{A_1, \dots, A_n, A\}$, siendo $A \in \Lambda$. Entonces se verifica:

$$A_t(A, G_{n+1}) - A_t(A, G_n) = A_t(A_{n+1}, G'_{n+1}) - A_t(A_{n+1}, G_n).$$

renombrando los elementos de los grupos en la forma

$$\forall i=1, \dots, n, \quad B_i \equiv A_i \quad \text{y} \quad B_{n+1} \equiv A$$

y aplicando el Teorema D se obtiene la expresión:

$$A_t(A_{n+1}, G'_{n+1}) - A_t(A_{n+1}, G_n) = F(B_{n+1} \cap A_{n+1}) - \sum_1^n F(B_i \cap B_{n+1} \cap A_{n+1}) + \quad /$$

$$+ \sum_{i < j}^n F(B_i \cap B_j \cap B_{n+1} \cap A_{n+1}) - \dots \pm \sum_{i < \dots < h}^n F(B_i \cap \dots \cap B_h \cap B_{n+1} \cap A_{n+1}) \pm \quad /$$

$$\pm F\left(\left(\bigcap_1^{n+1} B_i\right) \cap A_{n+1}\right). \text{ Deshaciendo el cambio de nombres de los elementos/}$$

y aplicando la conmutatividad de la intersección se obtiene el resultado:

$$A_t(A_{n+1}, G'_{n+1}) - A_t(A_{n+1}, G_n) = F(A_{n+1} \cap A) - \sum_1^n F(A_i \cap A_{n+1} \cap A) + \quad /$$

$$+ \sum_{i < j}^n F(A_i \cap A_j \cap A_{n+1} \cap A) - \dots \pm \sum_{i < \dots < h}^n F(A_i \cap \dots \cap A_h \cap A_{n+1} \cap A) \pm F\left(\left(\bigcap_1^{n+1} A_i\right) \cap A\right), \quad)$$

expresión que coincide, por el Teorema D, con $A_t(A, G_{n+1}) - A_t(A, G_n)$.

El resultado del corolario muestra la invariancia de la afinidad/ al intercambiar un elemento en un grupo y el resultado que produce induce a definir la afinidad entre un grupo y la intersección de dos elementos de acuerdo a:

1.7 Afinidad grupo-intersección de dos elementos.

Sean el grupo $G = \{A_1, A_2, \dots, A_n\}$ y los elementos $B, C \in \Lambda$. La afinidad entre G y $B \cap C$ se define como:

$$A_t(B \cap C, G) = \sum_1^n F(A_i \cap (B \cap C)) - \sum_{i < j}^n F(A_i \cap A_j \cap (B \cap C)) + \dots \pm \quad /$$

$$\pm \sum_{i < \dots < h}^n F((A_i \cap \dots \cap A_h) \cap (B \cap C)) \pm F\left(\left(\bigcap_1^n A_i\right) \cap (B \cap C)\right).$$

Como consecuencia de la definición combinada con los teoremas y / corolarios anteriores se obtienen los resultados:

1) $A_t(B \cap C, G) = F(B \cap C) + A_t(B, G) - A_t(B, G')$, siendo $G' = G \cup C$.

es trivial por el Teorema D aplicado a $A_t(B, G)$ y $A_t(B, G')$.

La elección de B ó C para construir G' es indiferente ya que, por el Corolario D2, si $G'' = G \cup B$ se tiene que /

$$A_t(B, G) - A_t(B, G') = A_t(C, G) - A_t(C, G'') \text{ y entonces} \quad /$$

$$A_t(B \cap C, G) = F(B \cap C) + A_t(C, G) - A_t(C, G'').$$

2) $A_t(B \cap C, G) = F(G \cap (B \cap C))$.

por definición de atracción.

3) El Teorema D puede expresarse en función de la atracción entre / G y $A_{n+1} \cap A$ según:

$$A_t(A, G_{n+1}) - A_t(A, G_n) = F(A_{n+1} \cap A) - A_t(A_{n+1} \cap A, G_n)$$

Una formulación equivalente es:

$$A_t(A, G_{n+1}) = A_t(A, G_n) + F(A_{n+1} \cap A) - A_t(A_{n+1} \cap A, G_n), \text{ manifes-}$$

tando una estructura equivalente a la del Teorema B.

1.8 Propiedades de Pal-Majumder.

Toda relación de afinidad elemento-grupo debe ser tal que verifique las relaciones intuitivas formuladas por Pal y Majumder ((4), /

pp. 88) aunque fueron establecidas anteriormente por Backer y Jain / ((9), pp.67-68):

- (i) La afinidad entre un elemento y un grupo no debe ser menor si el elemento es un miembro del grupo que si no está contenido/ en el grupo.
- (ii) La afinidad será aproximadamente 0 si el elemento es muy ex-/ traño respecto al grupo ("si el elemento está distante del / grupo ó fuera de la región de interés" sic.).
- (iii) La afinidad será igual a un máximo absoluto si el grupo con-/ siste en un único elemento que tenga la misma localización / que el elemento bajo consideración.

Estos conceptos intuitivos también han sido incorporados en la de finición de funciones de pertenencia para conjuntos difusos, fundamen talmente en problemas de reconocimiento dirigido (Pal y Majumder, (6) y en funciones de pertenencia de elementos a cluster para los conjun tos difusos inducidos por una partición (Backer y Jain,(9), pp. 68).

El concepto de afinidad definido en 1.1 verifica estas especifica ciones de acuerdo a:

- (i) por el Corolario D1.
- (ii) por las propiedades 1) y 2) de la afinidad.
- (iii) el Corolario D1 asegura que $F(A)$ es un máximo absoluto para / el elemento A, independientemente del grupo considerado. Sin/ embargo, dicho máximo no es absoluto para la distribución y / varía según el elemento A en consideración.

La noción de afinidad elemento-grupo sugiere, de forma natural, / la extensión a una definición de atracción ó afinidad grupo-grupo que permita reflejar la estructura común existente entre ambos grupos (si existe) y las relaciones que los enlazan, en base a las afinidades es tablecidas entre los elementos de los grupos.

Esta afinidad grupo-grupo no será simplemente la adición de las a finidades existentes entre sus elementos puesto que en este caso pue de aparecer más de una vez algún valor de afinidad entre elementos.

2 AFINIDAD GRUPO-GRUPO.

2.1 Afinidad grupo-grupo.

Sean los grupos G y G' constituidos por elementos de Δ . La afinidad entre G y G' se define como

$$A_t(G, G') = F(G \cap G').$$

Por la distributividad de unión e intersección esta definición es equivalente a:

$$A_t(G, G') = F\left(\bigcup (X \cap Y)\right), \quad X \in G, Y \in G'.$$

Enumerando los elementos de G y G' según $G = \{A_1, A_2, \dots, A_{n_1}\}$ y $G' = \{B_1, B_2, \dots, B_{n_2}\}$, la atracción entre ambos toma la forma:

$$A_t(G, G') = F\left(\bigcup_{i,j} (A_i \cap B_j)\right), \quad i=1, \dots, n_1, \quad j=1, \dots, n_2.$$

Intuitivamente, $A_t(G, G')$ es una medida de la información común a los elementos de los grupos tomados dos a dos sin tomar en cuenta los valores repetidos en las coincidencias, computando las conexiones de cada elemento de G con todos los de G' (ó viceversa).

2.1 Propiedades de la afinidad grupo-grupo.

1) Si $\forall X \in G$ y $\forall Y \in G'$ se tiene que $X \cap Y = \emptyset$, entonces $A_t(G, G') = 0$.

trivial por definición. Este valor constituye un mínimo para la afinidad entre grupos y corresponde a grupos que no tienen ningún estado en común en ninguna variable.

2) $0 \leq A_t(G, G') \leq F(G)$ y $0 \leq A_t(G, G') \leq F(G')$.

por la propiedad 4) de la intersección se tiene que:

$$F(G \cap G') \leq F(G) \quad \text{y} \quad F(G \cap G') \leq F(G').$$

$$3) A_t(G, G') \leq \frac{F(G) + F(G')}{2}.$$

basta sumar ordenadamente las desigualdades de la propiedad 2) anterior.

$$4) A_t(G, G') \leq F(GUG').$$

por la anterior propiedad 2) y la 4) de la unión se tiene $0 \leq A_t(G, G') \leq F(G)$ y $F(G) \leq F(GUG')$.

$$5) \text{ Si } G' \subset G, \text{ entonces } A_t(G, G') = F(G').$$

al ser $G' \subset G$ se deduce que $G \cap G' = G'$, por lo que el campo de ambos coincide.

$$6) \text{ Si } G \equiv G' \text{ entonces } A_t(G, G') = F(G), \text{ y } F(G) \text{ constituye un máximo para la afinidad entre } G \text{ y cualquier grupo.}$$

puesto que $G \cap G' = G = G'$, se tiene que $A_t(G, G') = F(G) = F(G')$. Y dado otro grupo G'' , como $G'' \cap G \subset G$, se deduce que la desigualdad $F(G \cap G'') \leq F(G)$ es cierta.

$$7) A_t(G, G') = F(G) + F(G') - F(GUG').$$

trivial por el Teorema B ya que $A_t(G, G') = F(G \cap G')$.

Los valores 0 y $F(G)$ (ó $F(G')$ si $F(G) < F(G')$) constituyen el mínimo y el máximo, respectivamente, de la afinidad entre dos grupos G y G' , aunque el máximo con respecto al grupo G es $F(G)$.

La propiedad 6) establece que la afinidad entre dos grupos no sobrepasará el valor del campo conjunto formado por los elementos de ambos grupos (GUG'), en consonancia con la idea de que el campo de un conjunto formado por la unión de dos grupos contiene la información total del conjunto mientras que la afinidad entre los grupos que lo constituyen computa solamente las interrelaciones entre los elementos de los grupos.

La aplicación del Teorema B a la afinidad grupo-grupo (en forma análoga al Teorema C para la afinidad elemento-grupo) origina una complicación computacional bastante grande puesto que aparecen múltiples

intersecciones de los elementos de G y G' en gran número. Un método / de cómputo alternativo consiste en calcular $A_t(G, G')$ de acuerdo a una medida basada directamente en la frecuencia (probabilidad) de los estados que aparecen en la distribución de unidades, para facilitar su / implementación en un computador. Esta idea lleva a definir una fun- / ción de pertenencia de un estado a dos grupos similar a la medida S / establecida por Ito, Kodama y Toyoda ((5), pp. 112).

2.3 Función de pertenencia de un estado a dos grupos.

Sean los grupos G y G' , y sea $x_{kh} \in W_k$. La función de pertenencia / $f^k: W_k \longrightarrow \{0, 1\}$ de x_{kh} a G y G' se define como:

$$f_{G, G'}^k(x_{kh}) = \begin{cases} 1 & \text{si } \exists A \in G \text{ y } \exists B \in G' / m_k(A) = m_k(B) = x_{kh}. \\ 0 & \text{en los demás casos.} \end{cases}$$

La familia de funciones f^k (una función para cada variable x_k) ex- / presa la idea intuitiva de la aparición de un estado común a G y G' , / y su valor será 0 sólo en el caso de estados que no están en ambos / grupos.

En términos de información, $f_{G, G'}^k$ permite estudiar la afinidad e / xistente entre G y G' de acuerdo a:

- a) un estado que no sea común a G y G' no influirá en la afini- / dad.
- b) si un estado es común a varios elementos de G y G' solamente / aparecerá una única vez.
- c) es suficiente que un estado sea común a un elemento de cada / grupo para que aporte su información a la afinidad.

Mediante las funciones $f_{G, G'}^k$ se obtiene una nueva expresión de / $A_t(G, G')$:

$$A_t(G, G') = - \sum_{k=1}^v \frac{1}{n_k} \cdot \sum_{h=1}^{n_k} \lg p(x_{kh}) \cdot f_{G, G'}^k(x_{kh}), \quad x_{kh} \in \psi_k.$$

Restringiendo el grupo G' a un solo elemento B , la anterior expresión adquiere validez para la afinidad punto-grupo particularizando f^k al elemento B según:

$$f_{B, G}^k(x_{kh}) = \begin{cases} 1 & \text{si } \exists A \in G / m_k(A) = m_k(B) = x_{kh}. \\ 0 & \text{en los demás casos.} \end{cases}$$

Entonces se tiene la expresión de la afinidad:

$$A_t(B, G) = - \sum_{k=1}^v \frac{1}{n_k} \cdot \lg p_k(m(B)) \cdot f_{G, G'}^k(m_k(B)), \quad \text{ya que se ve-}$$

rifica que $p_k(m(B)) = p(m_k(B)) = x_{kh}$ y $f_{G, G'}^k(x_{kh}) = 0$ si $m_k(B) \neq x_{kh}$.

Estas formulaciones de A_t son más fáciles de computar y disminuyen considerablemente el tiempo de cálculo al eliminar las referencias a intersecciones múltiples del tipo de las que aparecen en el Teorema C de afinidad punto-grupo. Este Teorema exige calcular la información que aportan todas las intersecciones de todos los elementos del grupo en número:

$$\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n - 1, \quad \text{siendo } n \text{ el número}$$

de elementos de G . Claramente, el número de cálculos es muy elevado y crece exponencialmente con n . Sin embargo, el cómputo a realizar para obtener $f_{B, G}^k$ se reduce a revisar los elementos de G para comprobar si coinciden con B en algún estado, es decir, basta con realizar n operaciones de cálculo (naturalmente, en ambas situaciones hay que extender estas consideraciones a todas las variables en cada uno de sus estados).

En el caso de los grupos G y G' con número de elementos n_1 y n_2 ,

respectivamente, el número de operaciones a realizar siguiendo las / pautas del Teorema C (realmente un teorema paralelo para grupos a par- / tir de la definición $A_t(G, G') = F(\cup(X \cap Y))$ ($X \in G, Y \in G'$)) se trans- / forma en:

$$\binom{n_1 n_2}{1} + \binom{n_1 n_2}{2} + \dots + \binom{n_1 n_2}{n_1 n_2} = 2^{n_1 n_2} - 1, \text{ mientras que el/}$$

cómputo a partir de $f_{G, G'}^k$ exige $n_1 n_2$ operaciones (como en el caso de / afinidad elemento-grupo, estas consideraciones deben extenderse a ca- / da variable en cada uno de sus estados).

Otra consecuencia importante del uso de la familia de funciones / f^k es la disminución computacional en la obtención de los campos con- / juntos $F(BUG)$ y $F(GUG')$ puesto que /

$$F(BUG) = F(B) + F(G) - A_t(B, G) \quad (\text{por la propiedad 4) de } A_t(B, G))$$

y $F(GUG') = F(G) + F(G') - A_t(G, G')$ (por la propiedad 7) de la afi- / nidad grupo-grupo). Estas relaciones son muy importantes para calcu- / lar sin gran costo de computador el campo conjunto del grupo formado / por la agrupación de los anteriores, más complicados de obtener usan- / do directamente el Teorema B.

2.4 La afinidad grupo-grupo y las propiedades de Pal-Majumder.

Las propiedades enunciadas por Backer y Jain y Pal y Majumder, / aunque referidas a la afinidad elemento-grupo, pueden extenderse a la / afinidad entre grupos según:

- (i') La afinidad entre dos grupos no debe ser menor si un grupo está contenido en el otro grupo que si no está contenido / en él.
- (ii') La afinidad será aproximadamente 0 si los grupos son muy / extraños entre sí.
- (iii') La finidad será igual a un máximo absoluto si ambos grupos

tienen los mismos elementos.

La afinidad grupo-grupo $A_t(G, G')$ definida anteriormente verifica/ estas propiedades:

- (i') por las propiedades 1) y 2) de A_t puesto que si $G' \subseteq G$, entonces $A_t(G, G') = F(G')$ y, para cada G , $A_t(G, G') \leq F(G')$.
- (ii') por la propiedad 1).
- (iii') por las propiedades 6) y 2), dado el grupo G fijo el máximo es $F(G)$, pero es un máximo absoluto solamente para el grupo G y varía de acuerdo al grupo G en consideración.

Debido a la definición de afinidad como medida de asociación entre grupos (ó elemento-grupo), destacan las siguientes peculiaridades básicas:

- a) Por construcción, la afinidad es tanto mayor cuanto más pequeña es la frecuencia (probabilidad en términos probabilísticos) de los estados que constituyen los grupos. Este hecho tiene el inconveniente de que, en el caso límite de elementos A, B, C, D tales que $A \equiv B$ y $C \equiv D$ se verifica que $A_t(A, B) > A_t(C, D)$ si los elementos A y B son más raros que C y D puesto que, para A y B , los estados son muy poco frecuentes y, por tanto, proporcionan mayor información.
- b) El número de elementos de G y G' no aparece directamente en la formulación de $A_t(G, G')$ pero influye fuertemente en la afinidad ya que cuanto más numerosos sean los grupos mayor es la posibilidad de algún elemento de algún grupo tenga un estado en común con algún elemento del otro grupo. Este hecho prima a los grupos grandes sobre los pequeños aunque en los últimos la asociación pueda ser mayor que en los primeros.

Estas consideraciones inducen la búsqueda de una medida de asociación inter-grupos que modere la influencia decisiva del número de ele

mentos en los grupos y de los valores de las frecuencias muy grandes, al menos en los casos extremos antes citados.

2.5 Asociación entre grupos.

Las ideas de Teoría de Conjuntos inducen una medida de asociación semejante a la similaridad de Ito, Kodama y Toyoda ((5), pp. 112 referida a atributos no-independientes) de acuerdo a:

$$r(A_i, A_j) = \frac{F(A_i) + F(A_j) - F(A_i \cup A_j)}{F(A_i \cup A_j)}, \quad A_i, A_j \in \Delta.$$

Por el Teorema A esta medida equivale, en este contexto de información, a

$$r(A_i, A_j) = \frac{F(A_i \cap A_j)}{F(A_i \cup A_j)} \quad \text{ó} \quad r(A_i, A_j) = \frac{A_t(A_i, A_j)}{F(A_i \cup A_j)}, \quad \text{pero /}$$

su extensión a similaridad entre grupos admite dos posibilidades fundamentales, de acuerdo a la extensión de $A_i \cap A_j$ que se considere:

- extensión de $A_i \cap A_j$ considerada como intersección de todos los elementos de G con todos los elementos de G' (atracción fuerte).
- extensión de $A_i \cap A_j$ considerada como $A_t(A_i, A_j)$, es decir, intersecciones entre cada elemento de G y cada elemento de G' (atracción-afinidad).

La relativización conseguida al dividir por $F(A_i \cup A_j)$ permite una comparación efectiva entre las similaridades de los elementos puesto que, en caso contrario, la homogeneidad entre elementos "raros" estaría bastante más valorada (no debe ocurrir en los casos extremos) que la homogeneidad entre elementos menos "raros".

Estas consideraciones llevan a definir dos medidas básicas de similaridad inter-grupos.

3 SIMILARIDAD INTER-GRUPOS.

3.1 Similaridad fuerte.

Esta medida generaliza $F(A \cap B)$, $A, B \in \Delta$ al campo originado por la intersección de los elementos de los grupos, computando la información proporcionada por los estados de las variables comunes a todos los elementos de los grupos.

Sean los grupos $G_1 = \{A_1, A_2, \dots, A_{n_1}\}$ y $G_2 = \{B_1, B_2, \dots, B_{n_2}\}$. La intersección total de G_1 y G_2 está definida por /

$$G_1 \wedge G_2 = \left(\bigcap_1^{n_1} A_i \right) \cap \left(\bigcap_1^{n_2} B_j \right) = A_1 \cap A_2 \cap \dots \cap A_{n_1} \cap B_1 \cap B_2 \cap \dots \cap B_{n_2} \quad , \quad /$$

(la notación \wedge es necesaria para evitar confusiones con $G_1 \cap G_2$ que co-

rresponde a $\left(\bigcup_1^{n_1} A_i \right) \cap \left(\bigcup_1^{n_2} B_j \right)$).

La información aportada por $G_1 \wedge G_2$ será por tanto /

$$F(G_1 \wedge G_2) = F(A_1 \cap A_2 \cap \dots \cap A_{n_1} \cap B_1 \cap \dots \cap B_{n_2}).$$

Es evidente que para elementos $A, B \in \Delta$ se tiene que $A \wedge B = B \wedge A$ y que $F(A \wedge B) = F(A \cap B)$.

Por definición, $F(G_1 \wedge G_2)$ verifica todas las propiedades del campo de la intersección:

1) Si $G_1 \wedge G_2 = \emptyset$, entonces $F(G_1 \wedge G_2) = 0$.

por la propiedad 1) de la intersección.

2) Si $G_1 \equiv G_2$, entonces se verifica que $F(G_1 \wedge G_2) = F\left(\bigcap_1^{n_1} A_i\right)$.

la demostración es trivial.

$$3) F(G_1 \wedge G_2) = F(G_2 \wedge G_1).$$

por la conmutatividad de la intersección.

$$4) F(G_1 \wedge (G_2 \wedge G_3)) = F((G_1 \wedge G_2) \wedge G_3).$$

por la asociatividad de la intersección.

$$5) F(G_1 \wedge G_2) \leq F(X), \text{ siendo } X \in G_1 \cup G_2.$$

por definición y la propiedad 1) de unión e intersección

$$6) \text{ Si } G_2' = G_2 \cup B, \text{ entonces } F(G_1 \wedge G_2) \geq F(G_1 \wedge G_2').$$

por la propiedad 2) de unión e intersección.

$$7) F(G_1 \wedge G_2) \leq F(G_1) \text{ y } F(G_1 \wedge G_2) \leq F(G_2).$$

tomando $X \in G_1$ se tiene que $F(X) \leq F(G_1)$ por la propiedad / 4) de unión e intersección, y por la anterior 5) se tiene / que $F(G_1 \wedge G_2) \leq F(G_1)$.

Para G_2 se usa un razonamiento análogo.

$$8) F(G_1 \wedge G_2) \leq F(G_1 \cup G_2).$$

por la propiedad 6) de unión e intersección.

De estas propiedades se deducen algunas consecuencias importan- / tes:

- el campo $F(G_1 \wedge G_2)$ es decreciente al aumentar el número de elementos en los grupos (propiedad 5)). La igualdad se produce si el nuevo elemento posee los mismos estados que son / comunes a ambos grupos.
- el campo $F(G_1 \wedge G_2)$ tiene como cota superior a $F(G_1)$ (ó a / $F(G_2)$). Naturalmente, también a $F(G_1 \cup G_2)$.

La intersección total entre G_1 y G_2 permite definir la medida de similitud entre ambos como:

$$S_1(G_1, G_2) = \frac{F(G_1 \wedge G_2)}{F(G_1 \cup G_2)} .$$

De la misma forma se define la similitud fuerte elemento-grupo/ y elemento-elemento:

$$S_1(A, G) = \frac{F(A \wedge G)}{F(A \cup G)} \quad \text{y} \quad S_1(A, B) = \frac{F(A \wedge B)}{F(A \cup B)} .$$

3.2 Propiedades de la similitud fuerte S_1 .

De acuerdo a las anteriores propiedades y a las propiedades de la unión, esta medida de similitud posee las siguientes características:

I) $0 \leq S_1(G_1, G_2) \leq 1$.

por la propiedad 8) de 3.1 se tiene que $F(G_1 \wedge G_2) \leq F(G_1 \cup G_2)$

y, por consiguiente, $S_1(G_1, G_2) \leq 1$.

La igualdad ocurrirá si se verifica que /

$$F(G_1 \wedge G_2) = F(G_1 \cup G_2) = F(G_1) = F(G_2) = F(x), \forall x \in G_1 \cup G_2 .$$

El valor 1 es una cota superior para S_1 que, en las condiciones anteriores (todos los elementos de los grupos coinciden), se transforma en máximo.

II) La similitud S_1 es decreciente al aumentar el número de elementos en los grupos, es decir, si $G'_1 = G_1 \cup B$, entonces se tiene que $S_1(G'_1, G_2) \leq S_1(G_1, G_2)$, puesto que:

por 6) se verifica que $F(G'_1 \wedge G_2) \leq F(G_1 \wedge G_2)$ y por la pro

propiedad 5) de unión e intersección se cumple la desigualdad /
 $F(G'_1 \cup G_2) \geq F(G_1 \cup G_2)$. De aquí se deduce:

$$\frac{1}{F(G'_1 \cup G_2)} \leq \frac{1}{F(G_1 \cup G_2)}, \text{ y multiplicando ordenadamente por la /}$$

desigualdad anterior $\frac{F(G'_1 \cap G_2)}{F(G'_1 \cup G_2)} \leq \frac{F(G_1 \cap G_2)}{F(G_1 \cup G_2)}$, es decir, /

$$S_1(G'_1, G_2) \leq S_1(G_1, G_2).$$

3.3 La similaridad fuerte y las propiedades de Pal-Majumder.

La similaridad fuerte S_1 verifica, de forma estricta, las propiedades de Pal-Majumder extendidas a medidas de atracción grupo-grupo:

(i') Si $G' \subset G$ se cumple que $F(G' \cap G) = F(G')$ y $F(G \cup G') = F(G)$.

Dado otro grupo cualquiera G'' no contenido en G , por la propiedad 7) de 3.1 se verifican:

$$F(G \cap G'') \leq F(G) \quad \text{y} \quad F(G \cup G'') \geq F(G) \quad (\text{por la propiedad 5) del campo de la unión. Y como } F(G') \leq F(G), \text{ se deduce que /}$$

$$S_1(G, G'') \leq S_1(G, G').$$

(ii') trivial por I).

(iii') La medida S_1 posee un máximo absoluto de valor 1 (de acuerdo a I)), y este máximo lo alcanza si se verifica que /

$$F(G_1 \cup G_2) = F(G_1) = F(G_2) = F(X), \forall X \in G_1 \cup G_2. \text{ Evidentemente /}$$

puede ocurrir que $F(G_1) = F(G_2) < F(G_1 \cup G_2)$ y en este caso no se alcanza el máximo.

La similaridad fuerte mide las relaciones fuertes que existen en-

tre los grupos, exigiendo coocurrencias en todos los elementos de ambos.

3.4 Similaridad-afinidad.

Tomando como base la afinidad entre grupos y, por tanto, incluyendo todas las relaciones entre sus miembros, surge una medida de similaridad según /

$$S_2(G_1, G_2) = \frac{A_t(G_1, G_2)}{F(G_1 U G_2)}$$

que es una generalización directa de la medida de Ito, Kodama y Toyoda (aunque en otro contexto) ya que, por la propiedad 7) de $A_t(G, G')$ / se tiene que /

$$S_2(G_1, G_2) = \frac{F(G_1) + F(G_2) - F(G_1 U G_2)}{F(G_1 U G_2)} .$$

Por definición se tiene que $S_2(G, G') = S_2(G', G)$ y mediante las propiedades de A_t se obtienen las siguientes propiedades de S_2 :

1) $0 \leq S_2(G, G') \leq 1$.

por la propiedad 4) de la afinidad entre grupos.

2) $S_2(G, G') = \frac{2 \cdot \bar{F}(G, G')}{F(G U G')} - 1$.

de la definición de S_2 se tiene que /

$$S_2(G, G') = \frac{F(G) + F(G')}{F(G U G')} - 1, \text{ y como } \bar{F}(G, G') = \frac{F(G) + F(G')}{2} \text{ se /}$$

obtiene de inmediato el resultado.

3) $S_2(G, G') \leq \frac{F(G)}{F(G U G')}$ y $S_2(G, G') \leq \frac{F(G')}{F(G U G')}$.

por la propiedad 2) de A_t .

$$4) S_2(G, G') \leq \frac{F(G) + F(G')}{2 \cdot F(GUG')} .$$

por la propiedad 3) de la afinidad puesto que /

$$A_t(G, G') \leq \frac{F(G) + F(G')}{2} .$$

$$5) \text{ Si } G \equiv G' \text{ entonces } S_2(G, G') = 1 .$$

por la propiedad 6) de A_t se tiene que $A_t(G, G') = F(G) /$
al ser $G \equiv G'$. Y como $GUG' = G$ entonces $F(GUG') = F(G)$.

3.5 La similaridad-afinidad y las propiedades de Pal-Majumder.

La similaridad-afinidad S_2 no verifica de forma estricta las propiedades de Pal-Majumder, fundamentalmente porque cualquier coocurrencia, por muy débil que sea, añade su efecto a la afinidad.

(i') Si $G' \subset G$, por la propiedad 5) de la afinidad se tiene que/
 $A_t(G, G') = F(G')$ y como $GUG' = G$ entonces $F(GUG') = F(G)$, de
donde se deduce que $S_2(G, G') = \frac{F(G')}{F(G)} .$

Dado otro grupo G'' no contenido en G cumple $F(G G'') \geq F(G)$
puesto que $G \subset GUG''$, pero no se verifica que /

$A_t(G, G'') \leq A_t(G, G') = F(G')$ en todos los casos. Solamen-
te será $S_2(G, G'') \leq S_2(G, G')$ si es válida la desigualdad /

$$\frac{A_t(G, G'')}{F(GUG'')} \leq \frac{F(G')}{F(G)} .$$

(ii') Se verifica por la propiedad 1) de S_2 .

(iii') Trivial por la anterior 1) de S_2 .

CAPITULO V

El objetivo fundamental del Análisis Cluster consiste en encontrar una partición de un conjunto de datos mediante una colección de conjuntos y bien separados denominados clusters. Cada elemento de un cluster debe verificar que su similaridad con cualquier miembro del cluster sea la mayor posible y con respecto a los demás clusters sea lo más disimilar posible. Se han desarrollado varios algoritmos y técnicas para obtener clusters de un conjunto de datos (un estudio detallado de técnicas y algoritmos aparece en Sneath y Sokal (15), Duda y Hart (19), Diday y Simon (18) y Escudero (16)).

1. LA ENTROPIA DE SHANNON Y EL CLUSTERING.

La teoría de la información ha contribuido al estudio de coeficientes de similaridad y métodos de agrupación entre variables cualitativas multiestado utilizando, fundamentalmente, el concepto de entropía de Shannon-Wiener por su carácter aditivo.

Una primitiva formulación de un coeficiente de similaridad fundado en la probabilidad y la información lo constituye la distancia de Rogers-Tanimoto definida para la unidad j como:

$$H_j = \sum_{k=1}^t d_{jk} = \sum_{k=1}^t (-\lg S_{jk}), \quad k \neq j, \quad /$$

siendo S_{jk} un índice de asociación entre las unidades j y k (S_{jk} indica cualquier coeficiente de asociación como: coeficiente de Jacquard, coeficiente simple matching, coeficiente de Yule, Coeficiente de Rogers-Tanimoto, coeficiente de Gower, etc.).

El uso de la entropía como fundamento del índice de similaridad ha sido llevada a cabo, inicialmente, por Goodall, Orloci y Estabrook definiendo la información (entendida como desorden, variación, etc.) para la variable i según:

$$H(i) = - \sum_{g=1}^{m_i} p_{ig} \cdot \ln p_{ig}, \quad /$$

siendo m_i el número de estados existentes en la variable i y p_{ig} la frecuencia relativa de las unidades que exhiben el estado g en la variable i . La información total de un grupo es, entonces, /

$$I = t \cdot \sum_{i=1}^n H(i), \quad /$$

siendo t el número de elementos del grupo y n el número de variables.

Los conceptos de información unión $I(h,i)$ e información mutua $I(h;i)$ entre las variables h,i , definidos por Orlocci y Estabrook, verifican la relación: /

$$I(h) + I(i) = I(h,i) + I(h;i)$$

y permiten la construcción de clusters aplicando técnicas SAHN (métodos de clustering Secuenciales Aglomerativos Jerárquicos sin Solapamiento) de acuerdo a:

la unión de dos grupos (clusters) ó elementos ó elemento-grupo produce un incremento ΔI en la información total, por lo que el algoritmo de agrupación debe minimizar este incremento en la entropía / definido por /

$$\Delta I = I(A,B) - I(A) - I(B), \quad /$$

siendo A y B clusters (Sneath y Sokal, (15) pp. 140-145, 240-244).

Recientemente, otros autores han estudiado algoritmos de agrupación y clasificación basados en el mínimo aumento de la entropía para agrupar dos clusters al recorrer los nodos de un árbol de decisión, aplicándolo al reconocimiento de caracteres (Wang y Suen, (3), pp. / 406-412).

Estas técnicas están basadas en la entropía clásica de Shannon / que considera la entropía media $H(x)$ de las probabilidades de los estados de las variables, sin influir el número de estados, etc.

2. LA DISTORSION Y LA AGRUPACION.

Las medidas basadas en la extensión de coeficientes de variables/dicotómicas a variables multiestado tienen en cuenta solamente los acuerdos y desacuerdos en cada variable (p.e. el coeficiente de Rogers Tanimoto) pero no consideran la posibilidad de que un acuerdo en una variable con 9 estados pueda ser de distinta naturaleza que un acuerdo en un carácter (variable) con 3 estados.

Las técnicas fundadas en la entropía de Shannon permiten tomar en cuenta la probabilidad del estado en que se realiza la coocurrencia / pero, por construcción, generalizan detalles acerca del número de estados de las variables que ayudan a un mejor conocimiento del espacio de unidades y de su estructura, es decir, la entropía de Shannon/ tiene como características:

- la entropía de un grupo es la información media proporcionada por los acuerdos existentes entre los elementos del grupo.
- no toma en cuenta el tipo de estado en que se produce la coincidencia de unidades (si es en variable con muchos ó pocos estados, etc.).

2.1 La distorsión como fundamento de la similaridad.

La utilización de la distorsión como base para la construcción de medidas de similaridad tiene por objetivo considerar varios factores/ esenciales en la descripción de una distribución de unidades:

- a) Proporciona la información total que suministra una unidad y, mediante el campo de un grupo, la información completa que subyace en dicho grupo.
- b) Al estar basada en la probabilidad (frecuencia) de aparición/ de cada estado en cada variable consigue que la influencia de/ la abundancia ó escasez de los estados se vea reflejada en su/ totalidad.

- c) Por construcción, el número de estados en cada variable está representado por el término $1/n_k$, considerando más efectivo un acuerdo entre estados de variables con pocos estados que en variables con muchos estados.

Una medida de similaridad construida a partir de la distorsión / tendrá en cuenta no sólo que existe coocurrencia en una variable sino también el tipo de variable en que se produce, la frecuencia del estado en que coinciden y la información total (campo) que incluya tanto acuerdos como desacuerdos.

2.2 La afinidad como índice de agrupación.

La elección de una medida cualitativa de asociación que abarque / el máximo posible de las características de las unidades es fundamental para poder establecer un índice de similaridad (ó una distancia, / ya que son opuestos) que permita cuantificar la semejanza entre unidades. Al tratarse de variables nominales y de acuerdo a la construcción de D_{n_i} , el índice de asociación verificará:

- todos los estados de una variable se consideran "a priori" igualmente significativos. A posteriori, su aportación será diferente en función de la información que aporte cada estado.
- la coincidencia de dos unidades en una u otra variable tiene distinto valor según el número de estados de cada variable.
- no existe a priori adjudicación de pesos a los distintos estados. La distribución de datos adjudica un peso a cada estado / en la información total de acuerdo a las afirmaciones anteriores.

Pueden pensarse otras medidas que difieran en parte de alguna de las condiciones anteriores como la construida por Ben-Bassat y Zaidenberg ((17), pp. 206-207) en que todos los estados porveen la misma información independientemente de la variable en que estén ubicados.

Desde el punto de vista de la agrupación, las distintas medidas /

de afinidad pueden ser utilizadas para construir índices de similaridad que permitan clusterizar el espacio de unidades, pero cada una posee unas características distintas de acuerdo a su definición y conducen a clusters con alguna distinción particular.

1) Similaridad fuerte S_1

El índice S_1 , por construcción, es un índice de similaridad que mide la atracción fuerte existente entre grupos ó elementos ya que un estado aporta información a S_1 si y sólo si aparece en todos y cada uno de los elementos en consideración (S_1 está basada en $F(G_1 \wedge G_2)$). Y puesto que es una medida relativa al campo total del grupo formado por la unión de los anteriores, tiene en cuenta los acuerdos y desacuerdos existentes en los grupos, que están incluidos en el campo conjunto de G_1 y G_2 .

Tiene como características esenciales:

- $S_1(G_1, G_2) = 0$ si ningún estado coincide en todos los elementos. Así, la coincidencia de sólo dos elementos en un único estado no es significativa en el total de G_1 y G_2 y no aporta información a S_1 .
- es simétrica por construcción.
- la relativización respecto a $F(G_1 \cup G_2)$ evita que S_1 alcance valores no significativos pues, sin ella, tendría un peso excesivo una coincidencia en estados raros, mayor que otra en estados frecuentes, lo que lleva a un contrasentido en casos extremos pues:
 si $G_1 \equiv G_1'$ y $G_2 \equiv G_2'$, sería mayor la similaridad entre G_1 y G_1' que entre G_2 y G_2' si los elementos de G_1 tuvieran estados más raros que los elementos de G_2 , lo que sería inaceptable.
- $0 \leq S_1 \leq 1$. Sólo será 1 si $G_1 \equiv G_2$.
- la propiedad II) de S_1 asegura que dados G_1 y G_2 , la adscripción de un elemento ó un grupo a cualquiera de ellos /

disminuya la similaridad y, por tanto, S_1 permite realizar / un análisis cluster de unidades mediante técnicas SAHN (método Secuencial Aglomerativo Jerárquico sin Solapamiento) siguiendo una monotonía creciente respecto al decrecimiento de la similaridad.

2) Similaridad-afinidad S_2

La medida S_2 , por construcción, computa todas las coincidencias entre los elementos de G_1 y G_2 , ya sean comunes a todos / los elementos ó solamente a dos de ellos (uno de cada grupo), / es decir, recibe la aportación tanto de los estados muy abundantes en los grupos como de los muy escasos por lo que, al realizar el análisis cluster, se fusionarán los grupos aún cuando / las coincidencias no constituyan un núcleo compacto (A y B coinciden en x_{ih} , C y D coinciden en x_{jk} , etc.), lo que conduce, a veces, a resultados poco consistentes.

La relativización originada por $F(G_1 \cup G_2)$ modera la influencia de la coincidencia, en casos límite, en estados raros puesto / que también toma en cuenta los desacuerdos al estar incluidos / en el campo conjunto de G_1 y G_2 .

De acuerdo a la definición de S_2 , destacan las siguientes características:

- $S_2(G_1, G_2) = 0$ si para cada elemento de G_1 no existe ningún elemento de G_2 con algún estado en común, y viceversa.
- es simétrica por construcción.
- la relativización respecto a $F(G_1 \cup G_2)$ consigue evitar los valores poco significativos para casos extremos de G_1 y G_2 (igual que ocurría con S_1).
- $0 \leq S_2 \leq 1$. El valor extremo 1 se alcanza si $G_1 \equiv G_2$, pero / también puede alcanzarse si $A_t(G_1, G_2) = F(G_1 \cup G_2)$ y se verifica $G_1 \neq G_2$; para ello basta que las coincidencias entre G_1 y G_2 cubran todos los estados comprendidos en $G_1 \cup G_2$. Este hecho posibilita que la similaridad entre dos grupos /

sea máxima, pero sin que exista un núcleo fuerte de coincidencias (sólo coincidencias aisladas).

$$- \forall G_1, G_2, S_1(G_1, G_2) \leq S_2(G_1, G_2).$$

$$\text{ya que } F(G_1 \wedge G_2) \leq A_t(G_1, G_2).$$

- S_2 no asegura que, dados dos grupos, la adscripción de un elemento o grupo a cualquiera de ellos disminuya la similaridad pues:

sean los grupos G y G' y $X \in \Delta$. Construyendo $G'' = G' \cup X$ puede ocurrir que $S_2(G, G') < S_2(G, G'')$ pues como $G' \subset G''$ se verifica que $F(G \cup G') \leq F(G \cup G'')$ y $A_t(G, G') < A_t(G, G'')$. De aquí que, /

al realizar los cocientes $\frac{A_t(G, G')}{F(G \cup G')}$ y $\frac{A_t(G, G'')}{F(G \cup G'')}$, no se puede de

terminar exactamente la relación entre $S_2(G, G')$ y $S_2(G, G'')$ / (en cada caso particular puede ocurrir cualquiera de las desigualdades ó la igualdad).

Estas consideraciones impiden la posibilidad de realizar un análisis cluster jerárquico basado directamente en S_2 puesto que no está asegurada una disminución de la similaridad según se desciende en el nivel de agrupación.

2.3 Medidas de similaridad derivadas de S_1 y S_2 .

El intento de solucionar estos inconvenientes conduce a varias medidas derivadas de las anteriores, inspiradas en otras medidas definidas en distintos contextos, fundamentalmente entropía fuzzy, similaridad fuzzy, fuzzynes, reconocimiento de imágenes, etc.

Para evitar estos resultados surgen varias medidas que, fundamentalmente, penalizan los grupos muy numerosos en base a considerar similaridades medias, medias logarítmicas y elementos iguales:

1) A partir de S_1 se puede considerar una medida derivada S_3 relativizando con respecto a $\bar{F}(G_1, G_2)$. Con ello se consigue que los valores sean algo mayores con respecto a S_1 puesto que se cumple $\bar{F}(G_1, G_2) \leq F(G_1 \cup G_2)$ ya que será $\bar{F}(G_1, G_2) \leq F(G_1)$ ó $\bar{F}(G_1, G_2) \leq F(G_2)$, y los campos de ambos grupos son menores que el campo/conjunto.

La medida S_3 se define como:

$$S_3(G_1, G_2) = \frac{F(G_1 \wedge G_2)}{\bar{F}(G_1, G_2)}, \text{ siendo } \bar{F}(G_1, G_2) = \frac{F(G_1) + F(G_2)}{2}.$$

Por construcción, S_3 consiste en tomar la media de los campos/ de los grupos en vez del campo conjunto y, en general, conservará las propiedades de S_1 pero con valores mayores que ella. Sin embargo, el hecho de considerar la media de los campos deja sin efecto la influencia total del campo conjunto de los grupos originando que, si la diferencia entre $F(G_1)$ y $F(G_2)$ es muy grande se produzcan efectos distorsionadores en los valores de S_3 que no se corresponden con la realidad, aumentando ó disminuyendo / improcedentemente las similaridades reales existentes.

Debido a los inconvenientes de S_2 , fundamentalmente el que ni $S_2/$ ni $A_t(G, G')$ son monótonas decrecientes respecto al número de elementos (excepto en los casos extremos de coincidencia de los grupos), se construyen otras medidas para paliar este problema:

sean los grupos G y G' tales que el número de elementos de $G \cup G'$ es n y n_e el número de elementos idénticos en $G \cup G'$, estos es

$$n_e = \begin{cases} 1 & \text{si } \exists X, \exists Y \in G \cup G' / X \neq Y. \\ n & \text{si } \forall X, \forall Y \in G \cup G' / X \equiv Y. \end{cases}$$

Las medidas derivadas de S_2 que pueden construirse son, entre otras:

$$2) S_4(G, G') = \frac{\lg n_e \cdot A_t(G, G')}{\lg n \cdot F(G \cup G')}.$$

Esta medida se inspira en la definición de entropía de un con

junto difuso tal como está definida (sin el término $\lg n_e$) por Kaufmann ((4), pp. 61), que mide la ambigüedad de un conjunto / difuso aunque como señalan Pal y Majumder ((4), pp. 129) es "in tuitivamente inaceptable" (sic) para computar la ambigüedad de / una imagen.

El término $\lg n_e$ (no aparece en la definición de Kaufmann) se / introduce para evitar el hecho que si G y G' tienen elementos i / guales, la similaridad S sea menor cuánto más elementos tengan / G y G' ya que entonces $A(G, G') = F(GUG')$ y $S(G, G') = 1/\lg n$.

$$3) \quad S(G, G') = \frac{n_e \cdot A_t(G, G')}{n \cdot F(GUG')} .$$

Se inspira en la entropía de un conjunto fuzzy de acuerdo a / De Luca y Termini ((10)) que adaptan la entropía teórica basada / en la información a las funciones de pertenencia de la Estadís / tica difusa (también es usada por Pal y Majumder (4), pp. 61 / para medir la ambigüedad de una imagen).

La entropía de De Luca y Termini es una extensión del concep / to de información de Shannon para n fuentes binarias al concep / to de fuzzynes, pero con un significado completamente distin / to. Con aplicación a la extracción de información en tratamien / to de imágenes y en Termodinámica Estadística, Xie y Bedrosian / ((7), pp. 151-153) utilizan este índice para definir la incer / tidumbre de un conjunto difuso compuesta de la información de / Shannon debida a la aleatoriedad y de la entropía de De Luca y / Termini como cantidad media de información difusa de informa / ción originada por la fuzzynes de un conjunto difuso relativo / a su conjunto ordinario.

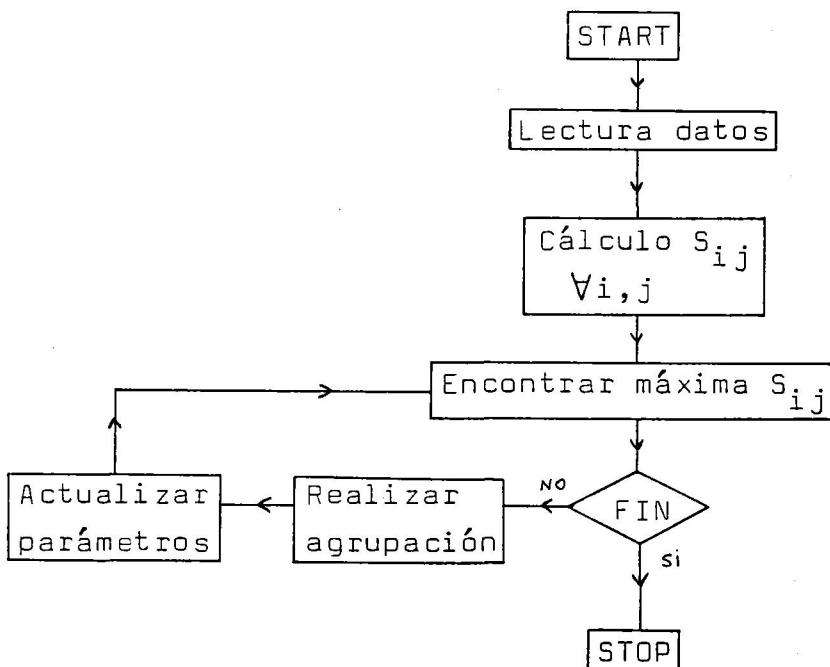
Pal y Chakraborty ((8), pp. 754-760) usan esta medida como un / índice de evaluación de patrones en términos de medidas inter / clases e intraclases, caracterizando y discriminando entre di / ferentes clases, con particular aplicación al reconocimiento / de vocales y de consonantes explosivas.

Numericamente, esta medida es semejante a la anterior en cuan / to a los valores que toma (aunque son más pequeños al cumplir / se que $\lg n < n$) y su significado es la media de la entropía t / tal aportada por un conjunto difuso (el término n_e no aparece / en la medida original).

3. ALGORITMO DE CLUSTERING.

El proceso de clustering consiste esencialmente en un módulo de / cálculo de parámetros necesarios (probabilidades, campos, similaridades) y un módulo de agrupación mediante un método jerárquico de maximización de la similaridad en cada nivel de agrupación que, en cada / paso de la jerarquía, agrupa los grupos ó elementos en un nuevo grupo que los engloba, realizado en base a varios procesos básicos:

- 1) Cálculo de probabilidades de los estados y de los campos individuales de las unidades.
- 2) Procedimiento de cálculo de similaridades intergrupos dos a / dos.
- 3) Algoritmo de agrupación y finalización.
- 4) Actualización de parámetros intergrupos en cada nivel de agrupación, de acuerdo a los resultados del paso anterior.



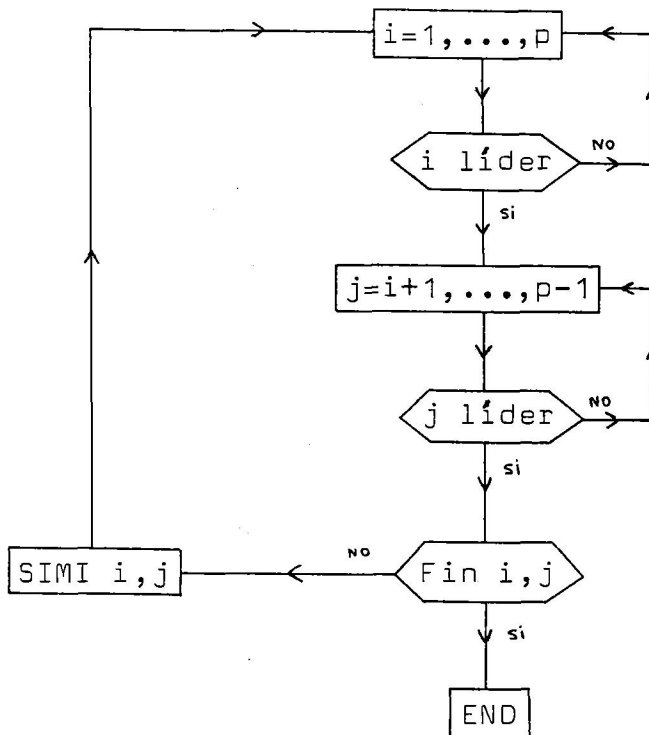
El algoritmo está fundamentado en varios procedimientos que realizan el cálculo de similaridades (CLAGRUP Y SIMI), un ciclo de agrupa-

ción (LOOP) y un procedimiento de actualización de parámetros (RECAL) junto a otros procesos secundarios (IGUAL Y NUMEL) que permiten trabajar con las medidas derivadas de S_1 y S_2 .

3.1 Procedimientos CALGRUP y SIMI.

Los procedimientos CALGRUP y SIMI actúan de forma anidada, siendo SIMI el procedimiento básico que es llamado repetidas veces ($p \cdot (p-1)$, siendo p el número de grupos en cada nivel de agrupación) por CALGRUP de acuerdo al desarrollo del proceso.

CLAGRUP localiza, entre todos los elementos, los líderes de los / grupos para, seleccionados dos a dos, calcular la similaridad entre / los grupos de los que son líderes mediante SIMI:

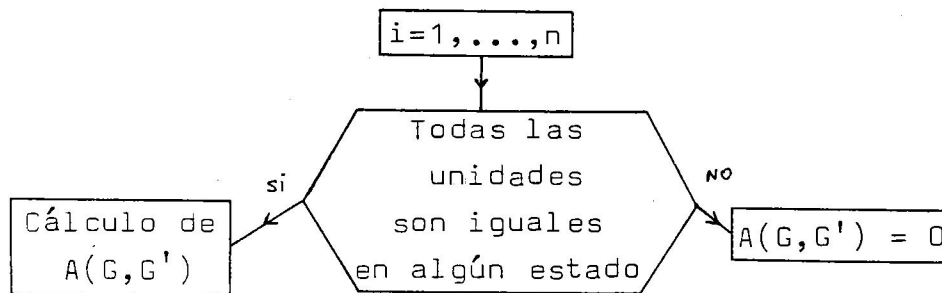


Una vez seleccionados los líderes de los grupos G y G' , su localización se transmite a SIMI como parámetros. SIMI marca los elementos de G y G' , calcula el número de elementos n de $G \cup G'$ y computa la afinidad $A_t(G, G')$ y el campo de $G \cup G'$ de acuerdo a la propiedad 7) de la afinidad grupo-grupo (ó la propiedad 4) de la afinidad elemento-grupo si uno de los grupos se reduce a un único elemento):

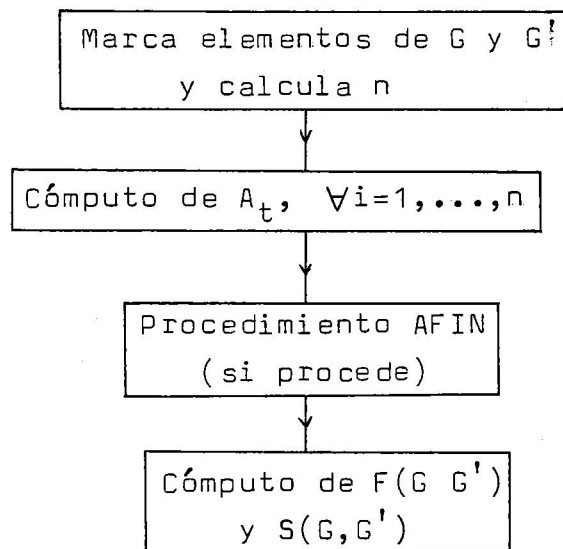
$$F(G \cup G') = F(G) + F(G') - A_t(G, G') \quad \text{ó} \quad /$$

$F(G \cup A) = F(G) + F(A) - A_t(A, G)$. En el caso de dos unidades, el Teorema A afirma que $F(A \cup B) = F(A) + F(B) - A_t(A, B)$.

En función de la medida de similaridad escogida, la base del índice de similaridad $A(G, G')$ será $A_t(G, G')$ si se escoge S_2 . Para S_1 se exige un nuevo procedimiento AFIN que calcule $F(G \wedge G')$ según:



SIMI finaliza calculando $S(G, G')$ en función de la atracción escogida y la similaridad elegida (S_1 , S_2 ó derivadas):

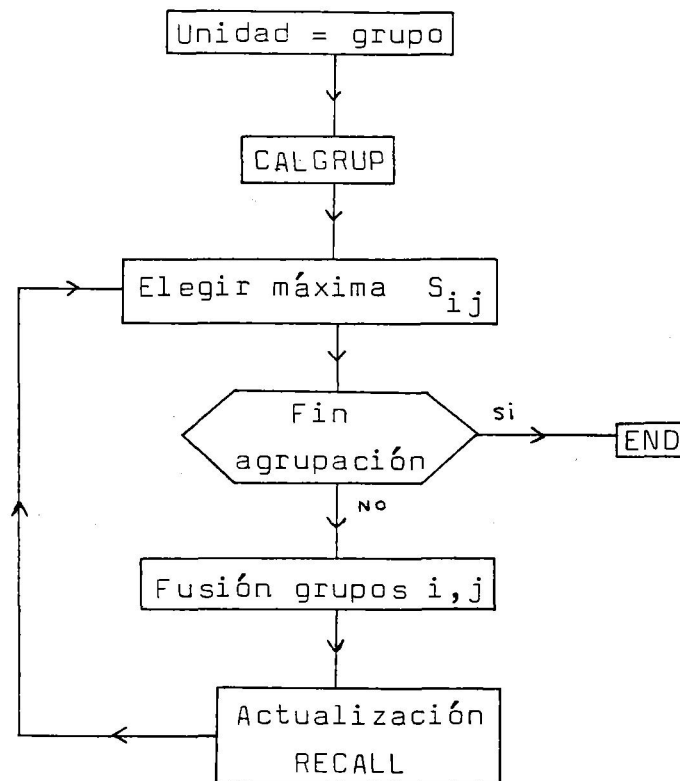


El procedimiento SIMI es fundamental a lo largo de todo el proceso de agrupación ya que constituye la base de otros procesos para computar la afinidad A_t , la atracción fuerte $F(G \wedge G')$ y el campo unión / de los grupos (provisionalmente si no se agrupan y definitivamente si lo hacen).

3.2 Procedimientos LOOP y RECALL.

El algoritmo de agrupación debe encontrar, en cada paso del proceso, los grupos con mayor similitud entre sí entendiéndose que un elemento no agrupado es también un grupo con un único miembro.

Mediante una iteración a lo largo de todos los líderes, que son / cabezas de grupo ó elementos no agrupados pero no miembros normales / del grupo, el proceso LOOP encuentra los líderes de los grupos con mayor similitud y, una vez localizados, procede a realizar la fusión/ de ambos grupos y la actualización de los parámetros necesarios (el / campo del nuevo grupo, las nuevas similitudes del mismo con el resto de grupos y la adscripción de los elementos de ambos en un sólo / grupo además de elegir el líder del mismo):

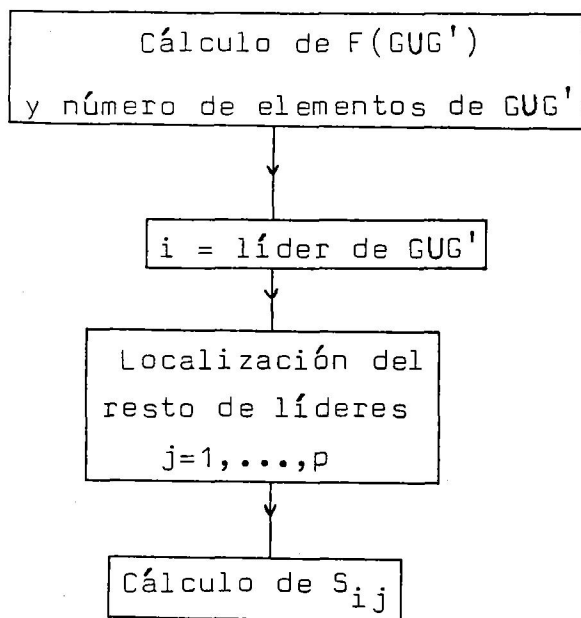


El final de LOOP se alcanza si todos los elementos están agrupa-/ dos en un único grupos que comprende toda la distribución ó si la má- xima similitud es 0, en cuyo caso se detiene el proceso con la exis- tencia de varios grupos claramente determinados. Otra posible forma / de finalización consiste en añadir el número de grupos deseado a prio- ri a las condiciones de fin de agrupación. Esta implementación es ele

mental introduciéndola como una condición más de fin de agrupación, aunque exige que, de acuerdo a algún criterio razonable, se pueda determinar a priori un número c de grupos adecuado.

El procedimiento RECAL es simple pero importante. Su misión consiste en:

en cada nivel de la agrupación, los grupos con mayor similitud G y G' deben agruparse y entonces hay que computar el campo del nuevo grupo G'' fusión de G y G' . Además, RECAL computa el número de elementos de G'' y las similitudes de G'' con el resto de los grupos mediante la identificación de los líderes de los mismos y de llamadas sucesivas a SIMI, de forma que se registre la modificación realizada en la distribución de grupos:



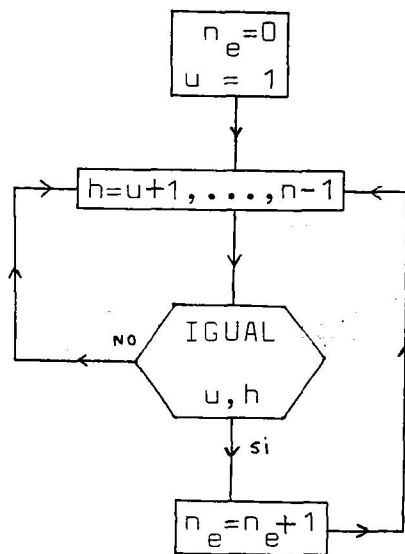
Las llamadas de LOOP a RECAL se realizan, como máximo, $p-1$ veces en cada una de las $p-1$ agrupaciones, siendo menor el número de llamadas cuanto mayor sea el nivel de la jerarquía.

3.3. Procedimientos auxiliares.

Los procesos CALGRUP, SIMI, LOOP y RECAL constituyen la base fundamental del algoritmo y, con ellos, se puede realizar una agrupación

en base a los índices S_1 y S_2 . Sin embargo, un análisis cluster realizado utilizando las medidas derivadas de ambas necesita otros procedimientos complementarios para cálculos menores: IGUAL y NUMEL.

IGUAL permite conocer si dos unidades coinciden ó no en todos sus estados, devolviendo el valor 1 si hay coincidencia y 0 si no la hay. Sucesivas llamadas a IGUAL permiten al procedimiento NUMEL calcular / cuantos elementos repetidos tiene GUG' y, si coincide con n , se obtiene el valor n_e necesario (en otro caso vale 1) para las medidas derivadas de S_1 y S_2 :



Una vez concluido el ciclo, si $n_e = n$ el número de elementos repetidos será n , pero si $n_e < n$ entonces el número de elementos repetidos es 0 ya que no coinciden todos los elementos.

Hay otros procesos menores que son necesarios aunque muy elementales:

- procedimiento que calcula las frecuencias de cada estado y los campos individuales de cada unidad.
- introducción de datos.
- escritura de datos en cada paso del cluster.

EJEMPLOS

EJEMPLO 1

CLASIFICACION DE MICROORDENADORES

Este ejemplo está extraído del trabajo de Michalski y Stepp de Julio de 1983 ((13), pp. 403-406), posteriormente analizado por Dale (/ (20), pp. 241-244) en Marzo de 1985, comparando los resultados del cluster conceptual con los de Taxonomía Numérica para llegar a la conclusión que ambas técnicas producen resultados semejantes.

La estructura de datos está constituida por un conjunto de 12 microordenadores definidos por las siguientes variables y estados:

	MP	RAM	ROM	Display	Teclas
1 Apple II	6502	48K	10K	Color	52
2 Atari 800	6502	48K	10K	Color	57-63
3 VIC 20	6502A	32K	11-16K	Color	64-73
4 Exidi Sor.	Z80	48K	4K	B/N	57-63
5 Zenith H8	8080A	64K	1K	Especial	64-73
6 Zenith H89	Z80	64K	8K	Especial	64-73
7 HP 85	HP	32K	80K	Especial	92
8 Horizon	Z80	64K	8K	Terminal	57-63
9 Ohio Chall.	6502	32K	10K	B/N	53-56
10 Ohio II	6502C	48K	10K	B/N	53-56
11 TRS-80 I	Z80	48K	12K	B/N	53-56
12 TRS-80 III	Z80	48K	14K	Especial	64-73

Evidentemente, las variables utilizadas para describir los datos/ no son significativas en algún caso (número de teclas y display); hay otros datos respecto a los microordenadores que serían más determinantes como: potencia del BASIC, capacidad en floppy-disk, OS implementada

do, etc. Además está bastante anticuada pero la idea consiste en poder utilizar una comparación tanto con el Cluster Conceptual como con la Taxonomía Numérica, por lo que seguirá la misma estructura.

Tal como señala Dale, la codificación de los datos realizada por Michalski y Stepp no está bien determinada ((20), pp. 241) apareciendo medidas dudosas como que la ROM está medida tanto en intervalos como en valores simples. Con estas reservas, la estructura realizada por Michalski y Stepp es:

MP { A: 6502, 6502A, 6502C.
B: Z80, 8080A.
C: HP.

RAM { A: 32K.
B: 48K.
C: 64K.

ROM { 0: 1K.
1: 4K.
2: 8K.
3: 10K.
4: 11-16K.
5: 12K.
6: 14K.
7: 80K.

Display { C: Color.
B: B/N.
E: Especial
T: Terminal.

Teclas { 1: 52.
2: 53-56.
3: 57-63.
4: 64-73.
5: 92.

PROBABILIDADES DE ESTADOS

ESTADOS

V A R	1	A .42	B .5	C .08								
	2	A .25	B .5	C .25								
	3	O .08	1 .08	2 .17	3 .33	4 .08	5 .08	6 .08	7 .08			
	4	C .25	B .33	E .33	T .08							
	5	1 .08	2 .25	3 .25	4 .33	5 .08						

Tabla 1.

INFORMACION DE ESTADOS

ENTROPIA DE SHANNON

ESTADOS

V A R	1	A 1.3	B 1	C 3.6								
	2	A 2	B 1	C 2								
	3	O 3.6	1 3.6	2 2.6	3 1.6	4 3.6	5 3.6	6 3.6	7 3.6			
	4	C 2	B 1.6	E 1.6	T 3.6							
	5	1 3.6	2 2	3 2	4 1.6	5 3.6						

Tabla 2.

INFORMACION DE ESTADOS

DISTORSION

ESTADOS

V A R	1	A .42	B .33	C 1.19								
	2	A .67	B .33	C .67								
	3	O .45	1 .45	2 .32	3 .2	4 .45	5 .45	6 .45	7 .45			
	4	C .5	B .4	E .4	T .9							
	5	1 .72	2 .4	3 .4	4 .32	5 .72						

Tabla 3.

En las tablas 2 y 3 se observa la diferencia entre la información de Shannon y la información producida por la distorsión, debido a la influencia del número de estados de las variables:

los estados x_{23} y x_{53} tienen la misma frecuencia (probabilidad) y por tanto proporcionan la misma información de Shannon $I = 2.$ Como las variables X_2 y X_5 tienen distinto número de estados, la información aportada por la distorsión es distinta, siendo mayor la del estado que pertenece a una variable con menos estados (X_2), obteniéndose $I(x_{23}) = .67$ e $I(x_{53}) = .40$ para la información producida por la distorsión.

ENTROPIA DE SHANNON. VARIABLES

1	2	3	4	5
1.33	1.5	2.75	1.86	2.13

Tabla 4.

DISTORSION. VARIABLES

1	2	3	4	5
1.95	1.67	3.21	2.19	2.55

Tabla 5.

El efecto de considerar información total y el número de estados/ aparece claro: la variable X_1 que en la tabla 4 tiene el mínimo valor pierde ese carácter en la tabla 5 debido a la distorsión que produce/ el estado x_{13} . La discriminación que pueda ocasionar X_1 es más rotunda que la producida por X_2 mediante la distorsión que si se considera la entropía de Shannon.

CAMPOS INDIVIDUALES
DE UNIDADES

1	2	3	4	5	6	7	8	9	10	11	12
2.17	1.85	2.35	1.91	2.16	2.04	3.42	2.62	2.08	1.75	1.91	1.83

Tabla 6.

La distorsión producida en el espacio de unidades por cada unidad individual está claramente de acuerdo con la abundancia de sus estados. Así el campo de las unidades 2 y 4 es bastante menor que el de la unidad 7 puesto que esta última posee estados en general bastante raros, sobre todo MP (es exclusivo de la unidad 7). Por tanto, la irregularidad creada por la unidad 7 produce una perturbación grande debido a la información que acarrean sus estados.

CAMPOS CONJUNTOS
DE GRUPOS

$$F(5,6) = 2.48, \quad F(1,2) = 2.57, \quad F(4,11) = 2.76 \quad F(9,10) = 2.4$$

$$F(5,6,12) = 3.27, \quad F(1,2,3) = 4, \quad F(9,10,4,11) = 4.04.$$

Tabla 7.

Se observa que el campo conjunto de varias unidades es mayor que el campo medio de las mismas ya que el primero computa los acuerdos y desacuerdos entre ellas mientras que el segundo computa los acuerdos y una media de la información producida por los desacuerdos. Naturalmente, cuanto mayores sean los campos individuales mayor será, en general, el campo conjunto dependiendo de los desacuerdos existentes entre sus componentes.

Los valores correspondientes a estas unidades y grupos de afinidad, intersección y similitudes se detallan a continuación en la tabla siguiente:

AFINIDADES

$$A_t(5,6)=1.71, \quad A_t(1,2)=1.44, \quad A_t(4,11)=1, \quad A_t(9,10)=1.42$$

$$A_t(8,(5,6))=1.32, \quad A_t(3,(9,10))=1.07, \quad A_t((1,2),(3,9,10))=1.44.$$

Tabla 8.

INTERSECCION FUERTE

$$F(5 \wedge 6)=1.71, \quad F(1 \wedge 2)=1.44, \quad F(4 \wedge 11)=1, \quad F(9 \wedge 10)=1.42$$

$$F(12 \wedge (5,6))=1.05, \quad F(3 \wedge (1,2))=.92, \quad F((1,2) \wedge (3,9,10))=.42.$$

Tabla 9.

Por definición, para unidades individuales se verifica que $F(A_i \wedge A_j) = A_t(A_i, A_j)$. La afinidad muestra su no decremento con la ampliación del grupo original puesto que al ampliar el grupo (5,6) con el elemento 8 si decrece pero la ampliación de (3,9,10) con (1,2) aumenta A_t

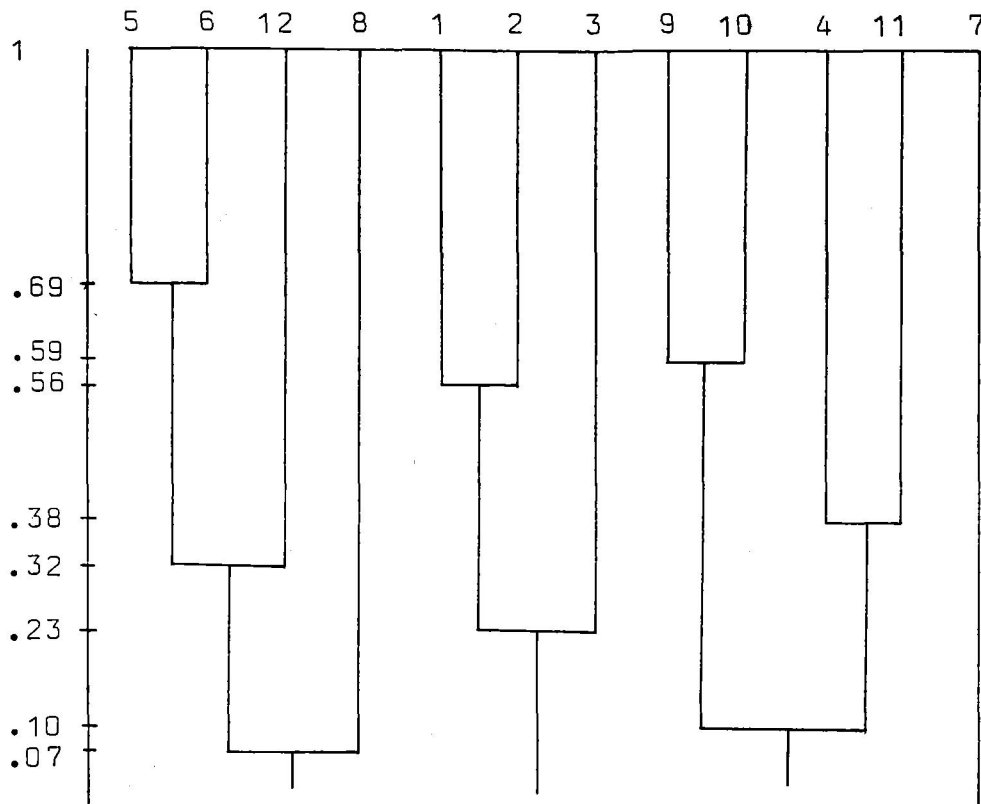
SIMILARIDADES

Grupos	S_1	S'_1	S_2	S_3	S_4
(5,6)	.69	.82	.69	.69	.34
(1,2)	.56	.72	.56	.56	.28
(9,10)	.59	.74	.59	.59	.29
((5,6),8)	.26	.39	.35	.22	.12
(3,(9,10))	.11	.18	.29	.19	.10
((1,2),(3,9,10))	.09	.13	.30	.13	.06

Tabla 10.

Se observa en S_2 el no decrecimiento de la similaridad al adjuntar (1,2) a (3,9,10).

AGRUPACION. DENDROGRAMAS



Indice S_1 . Tabla 11

Este mismo esquema aparece, sin escala, en la siguiente tabla para poder comparar los resultados con los de Michalski y Stepp.

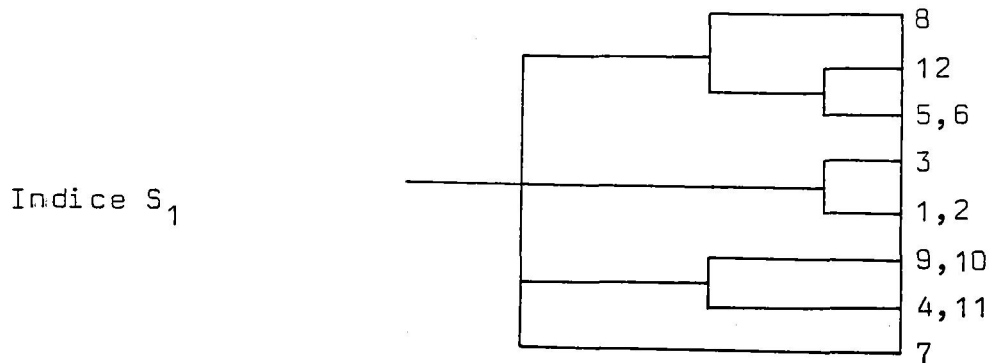
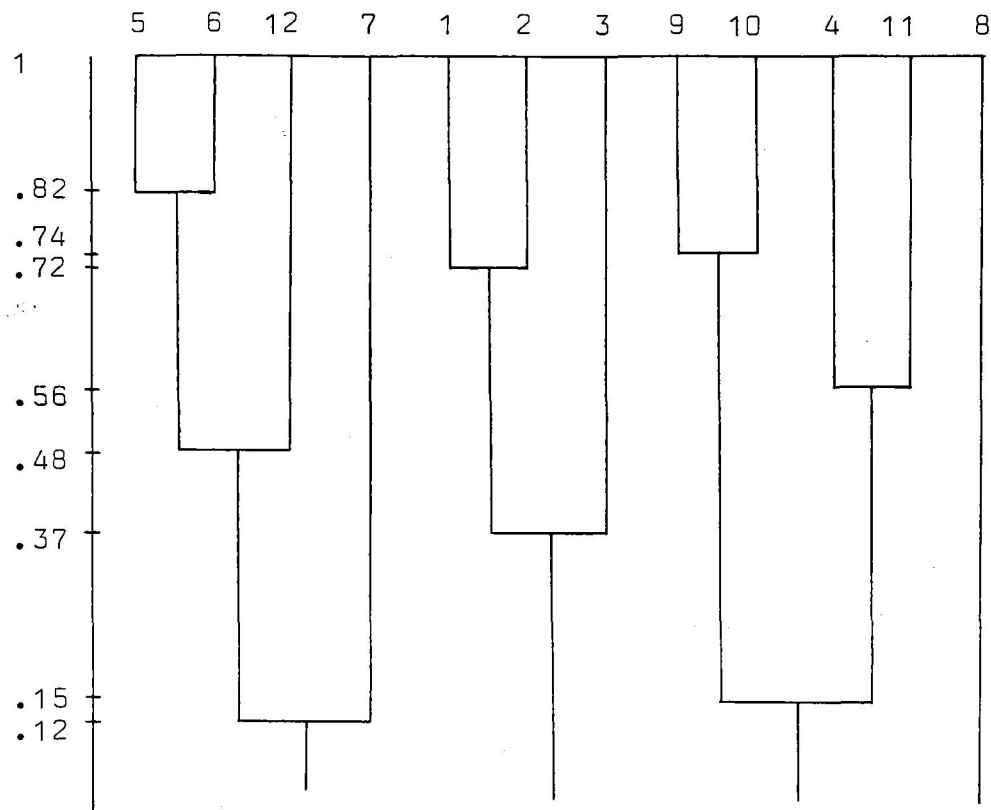


Tabla 12.



Indice S'_1 . Tabla 13.

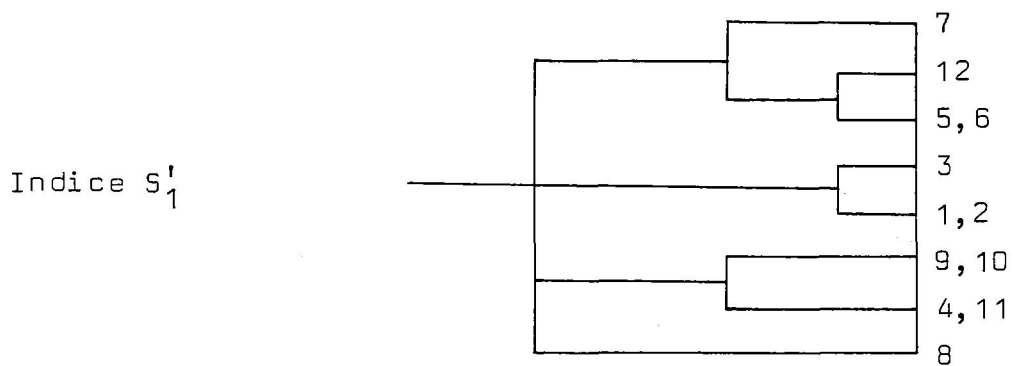
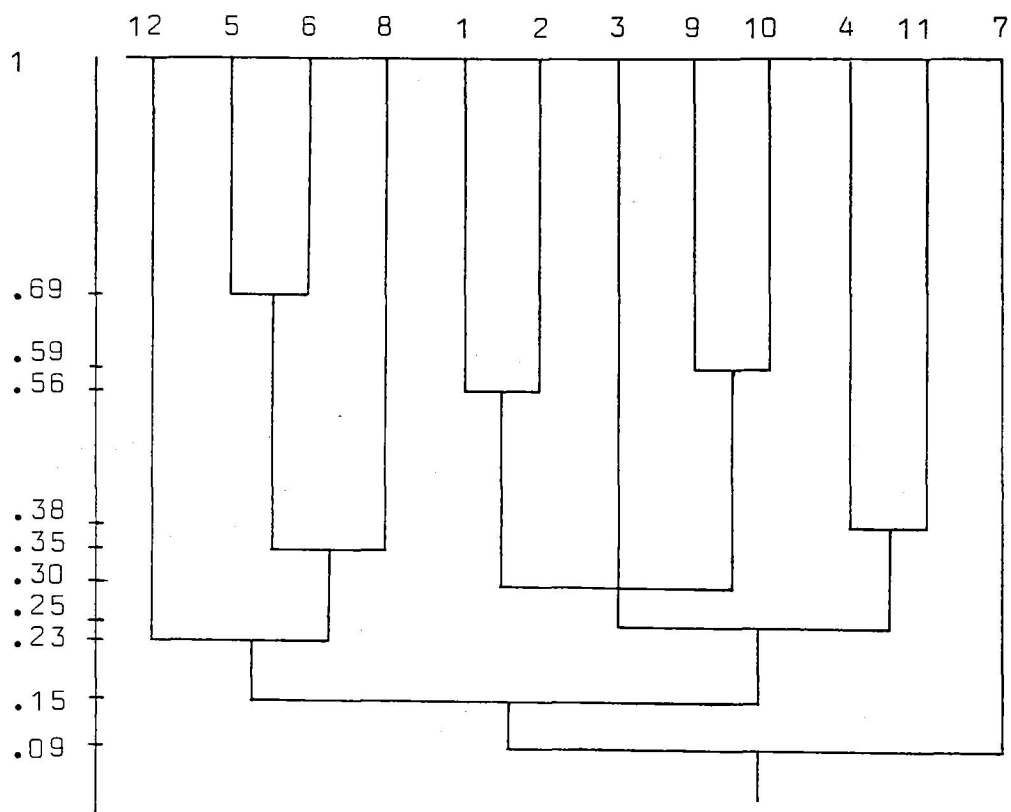


Tabla 14.

La diferencia entre ambos resultados está clara. Mientras S_1 aisla claramente a la unidad 7 por sus características espúreas, S'_1 la a grupe con (5,6),12 aislando a la unidad 8, debido a tomar la media de los campos de los grupos en vez del campo conjunto. Este hecho hace /

que los valores de S_1' sean mayores, en general, que los de S_1 aunque/
con el inconveniente antes mencionado.



Indice S_2 . Tabla 15.

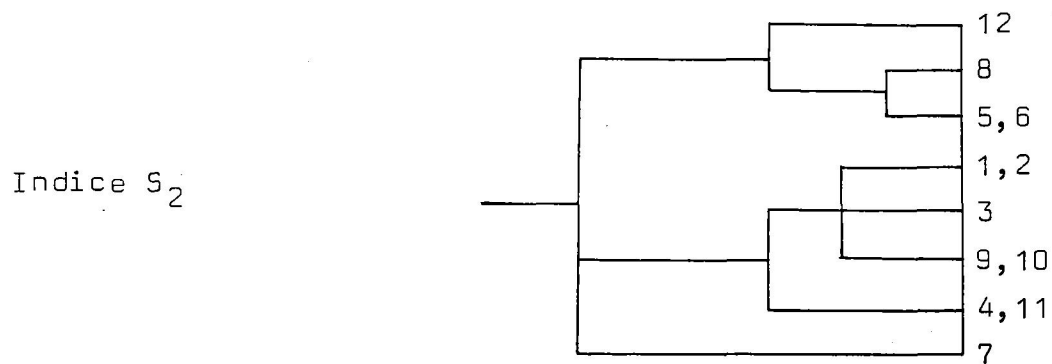
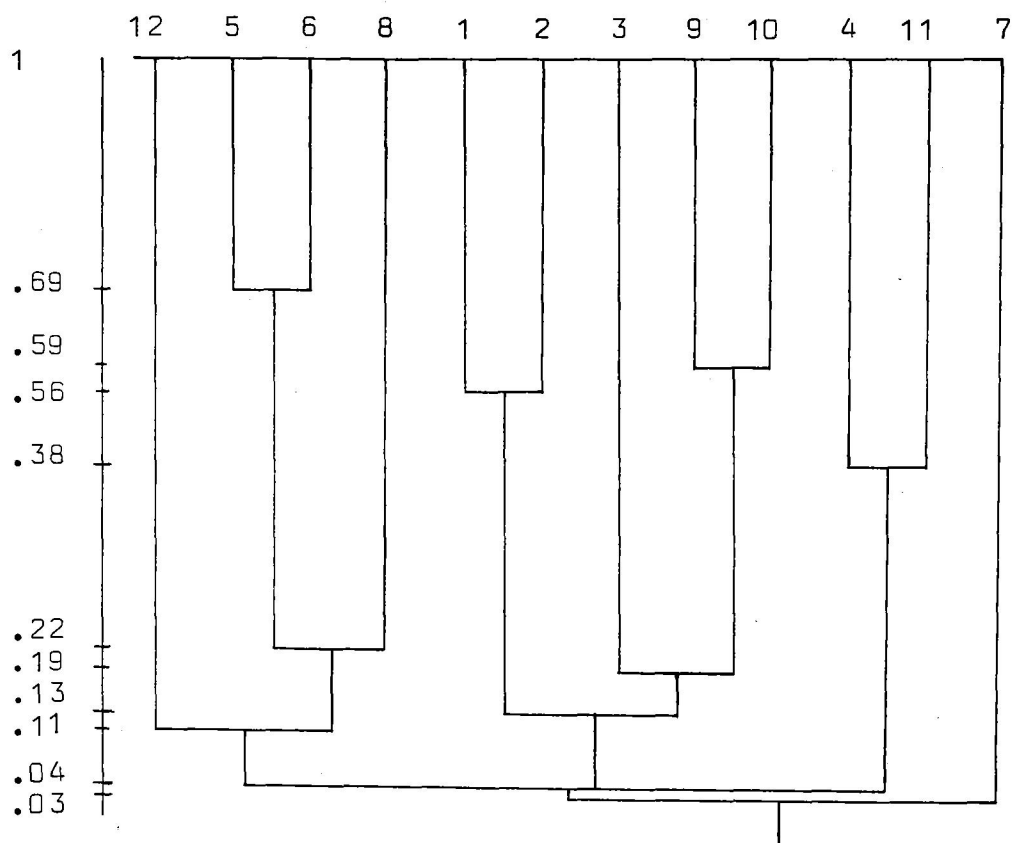


Tabla 16.

En el nivel donde aparece el grupo $((1,2),3,(9,10))$ surge el pro-

blema puesto que, originalmente se agrupan (9,10) y 3 con $S_2=.29$ y / en el siguiente nivel lo hacen (1,2) y ((9,10),3) con $S_2=.30$. Este hecho puede soslayarse en la forma anterior pero no hay seguridad de / que sea una ocurrencia aislada debido al no decrecimiento de S_2 en todas las situaciones.



Indice S_3 . Tabla 17.

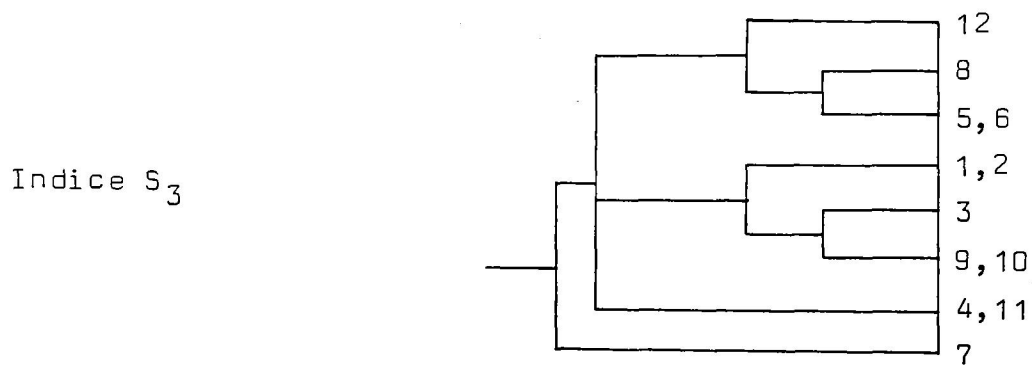
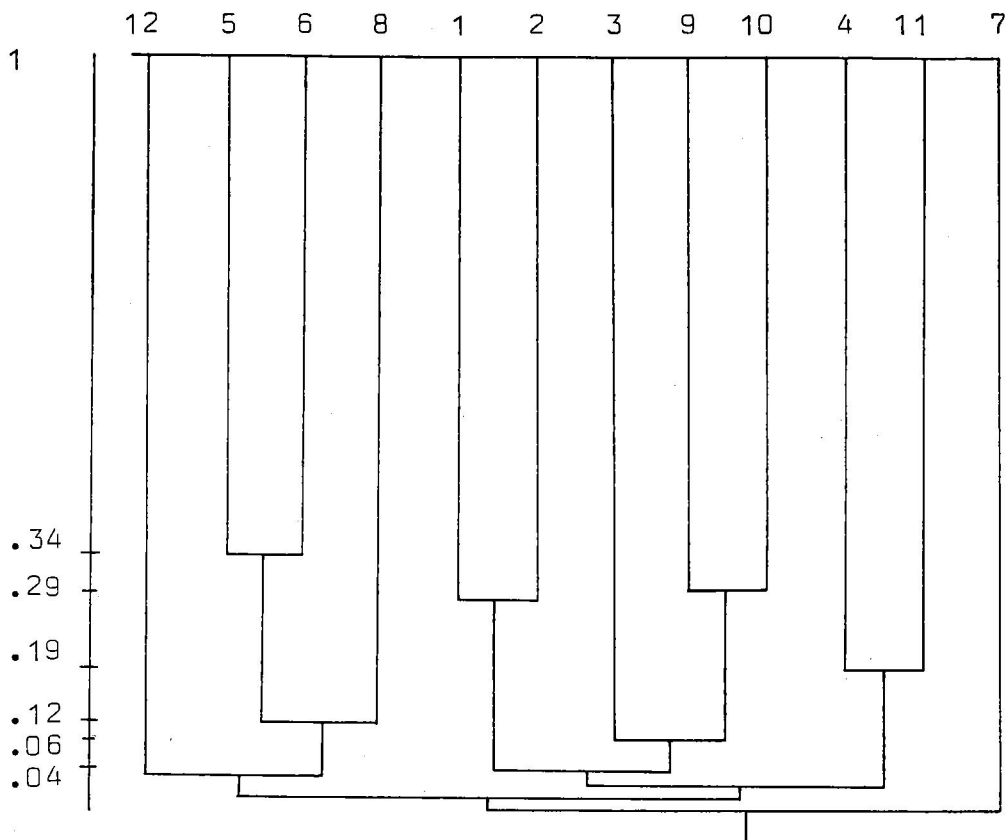


Tabla 18.

En este caso ocurre igual que con S_2 en $S_3=.04$ ya que, original-/

mente, se habian agrupado ((1,2),(3,9,10)) y (4,11) en primer lugar / para agruparse (1,2,3,9,10,4,11) y (12,5,6,8) en el siguiente nivel./ Como en el caso anterior se han englobado ambas agrupaciones en una / sola más amplia.



Indice S₄. Tabla 19.

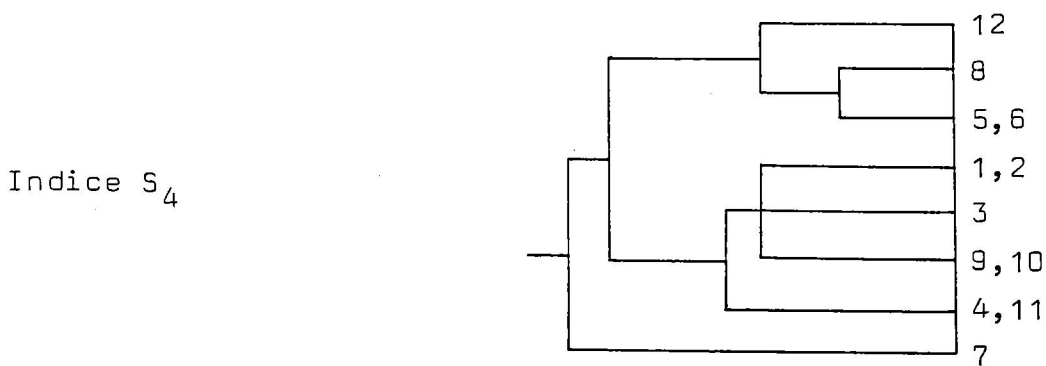


Tabla 20.

Este dendrograma origina varias dificultades puesto que, al ser va lores medios, las cantidades numéricas son muy pequeñas y bastante se

mejantes entre grupos, por lo que se han omitido las referencias a S_4 en algunos niveles del dendrograma que alcanzaban valores tales como/ .29 y .28, .06, .05 y .04. Sin embargo, no aparecen problemas debido/ al decrecimiento aunque no hay seguridad de que no ocurran en otros / casos (realmente, los valores tan próximos de S_4 inducen a pensarlo).

CLASIFICACION DE MICHALSKI Y STEPP

CLUSTER CONCEPTUAL

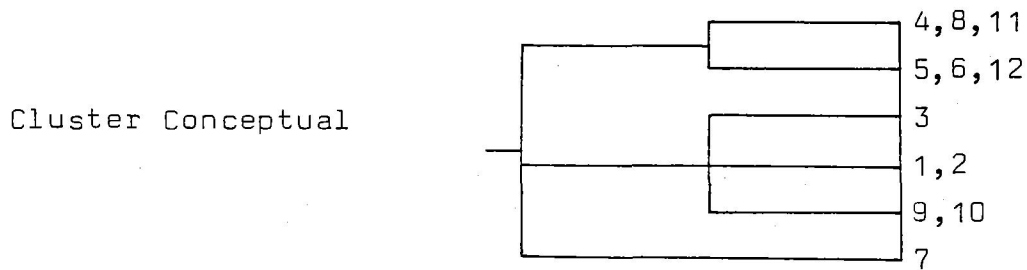


Tabla 21.

RESULTADOS DE TAXONOMIA NUMERICA (DALE)

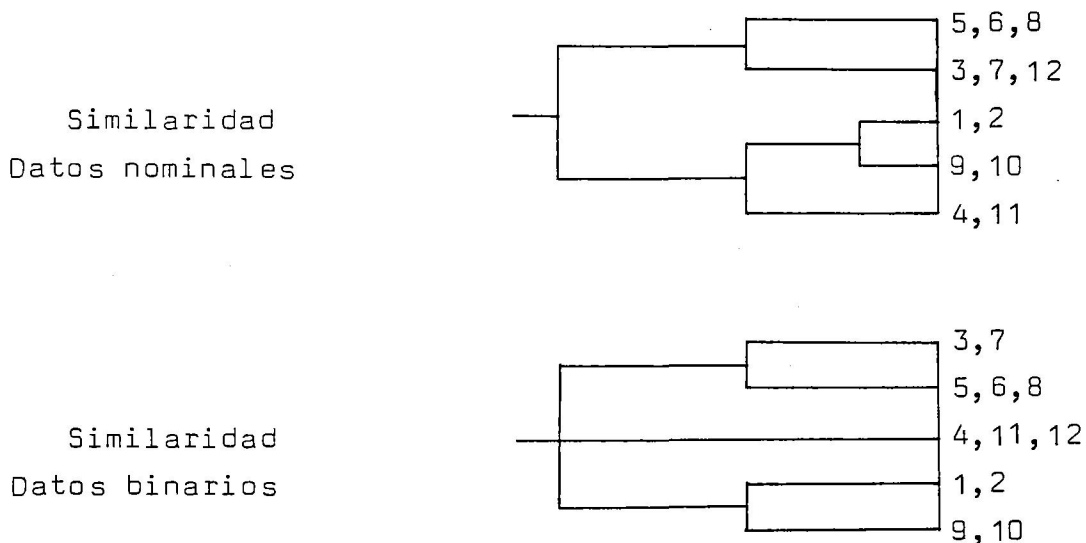


Tabla 22.

Claramente el cluster conceptual separa la unidad 7 de forma ro-/

tunda, mientras que los diagramas de la tabla 22 lo agrupa de forma/ bastante extraña (las variables del elemento 7 son muy distintas de / las de los demás elementos).

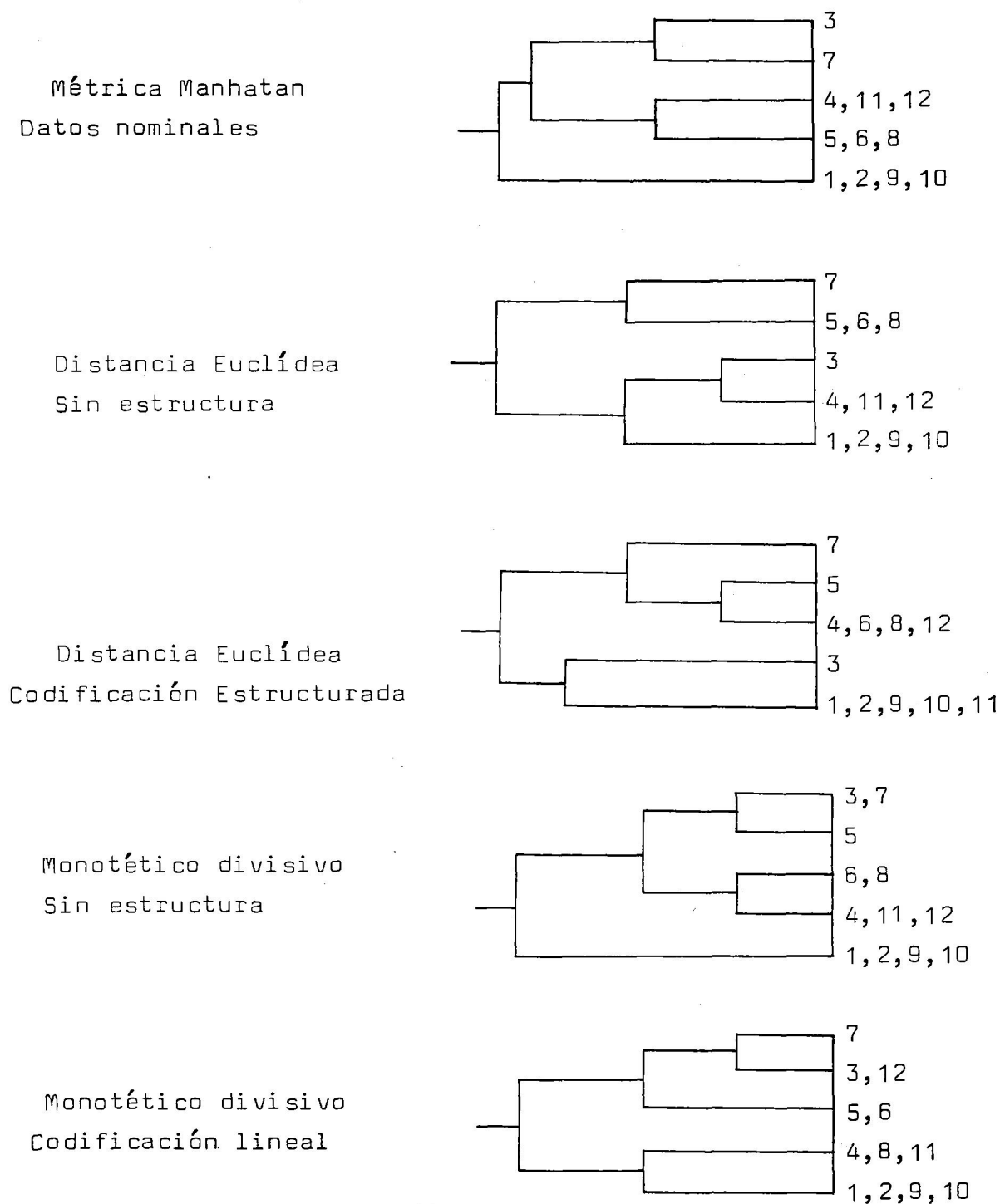


Tabla 23.

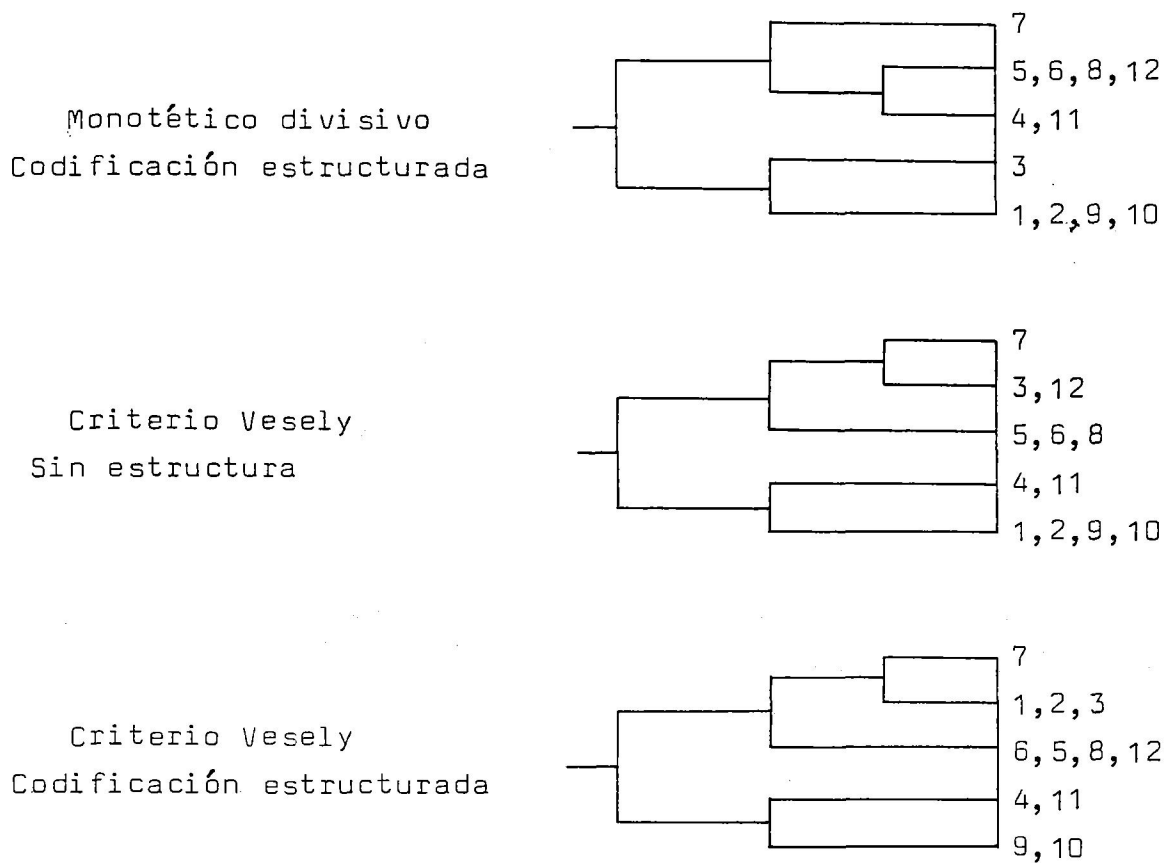


Tabla 24.

Los resultados de Dale son menos satisfactorios que los obtenidos por el cluster conceptual y destaca la estructuración sobre la distribución no estructurada, aunque no está demasiado claro como se consigue la estructura. Los resultados del cluster conceptual están más de acuerdo con los datos y resultan bastante parecidos a los obtenidos mediante el procedimiento monotético divisivo con codificación estructurada.

Los dendogramas obtenidos a partir de las similaridades basadas en la distorsión se dividen en dos grupos fundamentales:

- a partir de S_1 y S'_1 se obtienen 3 grupos diferenciados formados por (5,6,12,8), (1,2,3) y (9,10,4,11) con el elemento 7 totalmente aislado (excepto en S'_1 que deja aislado al elemento 8 lo que no ocurre en ninguno de los demás dendogramas). Estos resultados se muestran bastante de acuerdo con los datos.

- a partir de S_2 y sus medidas derivadas se observa la misma / composición: 3 grupos claramente diferenciados que son los / mismos en todos los casos (5,6,12,8), (1,2,3,9,10) y (4,11), / quedando nuevamente aislado el elemento 7. Comparados con los anteriores basados en S_1 queda de manifiesto que la homogenei- / dad es mayor ya que en el anterior se agrupaba (9,10) con el / grupo (4,11) y, en función de los valores de las similarida- / des, solamente la agrupación basada en S_4 muestra una estructu- / ra extraña debido a la escasa magnitud de las similaridades.

Los datos anteriores están extraídos de Michalski y Stepp (13) y / Dale (20), aunque en este último caso la exposición de resultados es / bastante sucinta sin realizar los dendogramas. En el primer caso se / ha utilizado el programa CLUSTER/2 mientras que en el segundo se ha u / tilizado el paquete de clasificación numérica TAXON en un ordenador / CYBER 76.

EJEMPLO 2

AGRUPACION DE CARACTERES ALFABETICOS

Sea el conjunto de letras constituido por 12 elementos que comprenden los caracteres A, B y b de acuerdo a la siguiente descripción

		Tamaño	Tipo	Color
1	A	M	A	R
2	A	M	A	R
3	A	M	A	V
4	A	M	A	Z
5	A	M	A	Z
6	A	M	A	Z
7	B	M	B	R
8	b	m	B	R
9	B	M	B	R
10	B	M	B	V
11	B	M	B	V
12	B	M	B	Z

Para su descripción se han tomado tres variables según:

X_1 = tamaño, X_2 = Tipo de letra y X_3 = color de acuerdo a la codificación:

$W_1 = \{M, m\}$, M = mayúscula, m = minúscula.

$W_2 = \{A, B\}$, A = letra a, B = letra b.

$W_3 = \{R, V, Z\}$, R = rojo, V = verde, Z = azul.

Claramente se observa que un carácter es muy escaso (b) y que la variable X_1 es la que representa menos homogeneidad.

PROBABILIDADES DE ESTADOS

ESTADOS

V A R	1	M .92	m .08		
	2	A .50	B .50		
	3	R .42	V .25	Z .33	

Tabla 1.

INFORMACION DE ESTADOS

ENTROPIA DE SHANNON

ESTADOS

V A R	1	M .13	m 3.58		
	2	A 1	B 1		
	3	R 1.26	V 2	Z 1.58	

Tabla 2.

INFORMACION DE ESTADOS

DISTORSION

ESTADOS

V A R	1	M .06	m 1.79		
	2	A .50	B .50		
	3	R .42	V .67	Z .53	

Tabla 3.

Las tablas 2 y 3 muestran la diferencia entre la información producida por la distorsión y la que proporciona la entropía de Shannon: el estado x_{22} aporta la mitad de la información que x_{32} en entro-

pía de Shannon mientras que no se mantiene la proporción en base a la distorsión ya que el número de estados en cada variable es distinto.

ENTROPIA DE SHANNON. VARIABLES

1	2	3
.41	1	1.55

Tabla 4.

DISTORSION. VARIABLES

1	2	3
1.85	1	1.6

Tabla 5.

Se observa claramente que la entropía de Shannon es, para cada variable, menor ó igual que la distorsión, de acuerdo con la teoría expuesta en el Capítulo I. Las relaciones de orden en cada una de ellas están y la variable con más distorsión es X_1 , mientras que la demayor entropía es X_3 notandose la influencia del término $1/n_k$.

CAMPOS INDIVIDUALES
DE UNIDADES

1	2	3	4	5	6	7	8	9	10	11	12
.98	.98	1.23	1.09	1.09	1.09	.98	2.71	.98	1.23	1.23	1.09

Tabla 6.

Se destaca el elemento 8 con un valor del campo de 2.71, como corresponde a la variable X_1 en su estado x_{12} (m).

CAMPOS CONJUNTOS
DE GRUPOS

$$\begin{aligned}
 F(1,2) &= .98, & F(4,5,6) &= 1.09, & F(10,11) &= 1,23 \\
 F(7,9) &= .98, & F(7,9,12) &= 1.51, & F(3,10,11) &= 1,73 \\
 & & F(1,2,4,5,6) &= 1.51. & &
 \end{aligned}$$

Tabla 7.

Comparando con la Tabla 6 se observa que los campos conjunto de / elementos iguales es el mismo que el campo individual de uno de los e lementos que lo integran puesto que la unión de ambos no gana información (la ganancia es igual que la información que había).

Los valores correspondientes de afinidades e intersecciones fuertes vienen dadas por:

$$\begin{aligned}
 A_t(1,2) &= .98, & A_t(4,5,6) &= 1.09, & A_t(10,11) &= 1.23 \\
 A_t(7,9) &= .98, & A_t(7,9,12) &= .56, & A_t(3,10,11) &= .73 \\
 & & A_t(1,2,4,5,6) &= .56. & &
 \end{aligned}$$

Tabla 8.

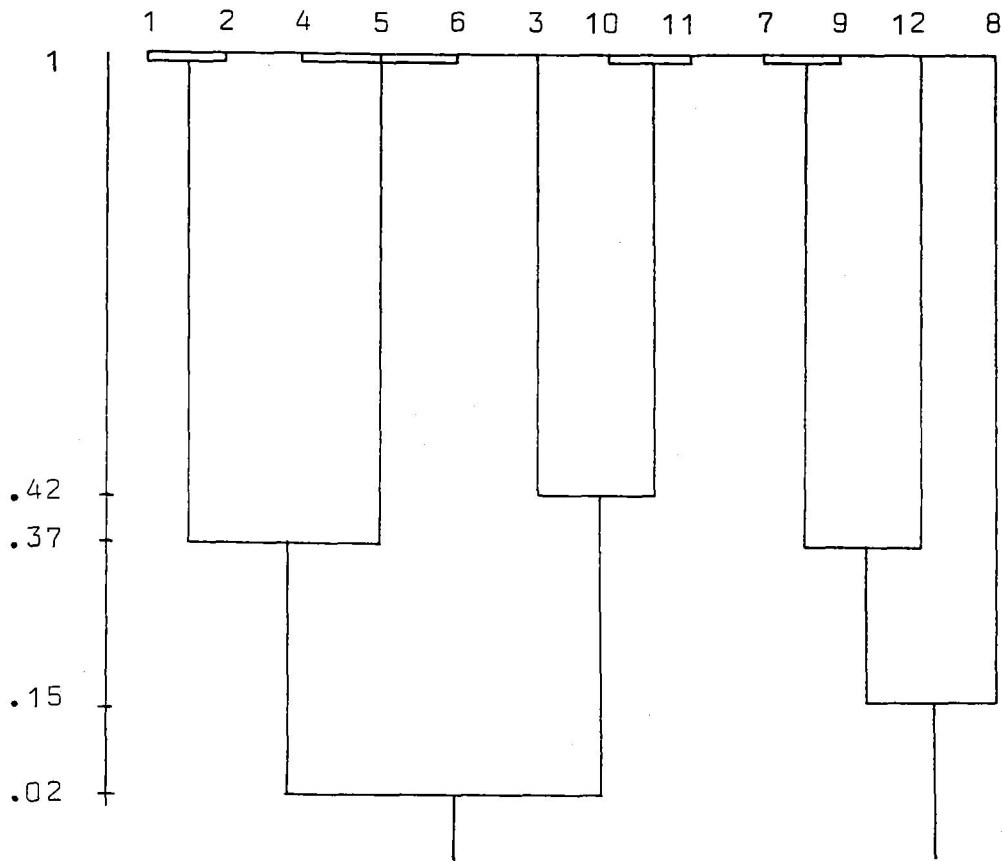
$$\begin{aligned}
 F(1 \wedge 2) &= .98, & F(4 \wedge 5 \wedge 6) &= 1.09, & F(10 \wedge 11) &= 1.23 \\
 F(7 \wedge 9) &= .98, & F(7 \wedge 9 \wedge 12) &= .56, & F(3 \wedge 10 \wedge 11) &= .73 \\
 & & F(1 \wedge 2 \wedge 4 \wedge 5 \wedge 6) &= .56. & &
 \end{aligned}$$

Tabla 9.

A primera vista parece que coinciden completamente afinidad e intersección, pero este caso solamente se da cuando es un grupo de sólo dos elementos ó sus elementos son unicamente de dos clases, tal y como aparece en la definición de intersección fuerte (Capítulo 4). Esta coincidencia no permite distinguir claramente el aumento de la afinidad en algunos casos con el número de elementos y el descenso parale-

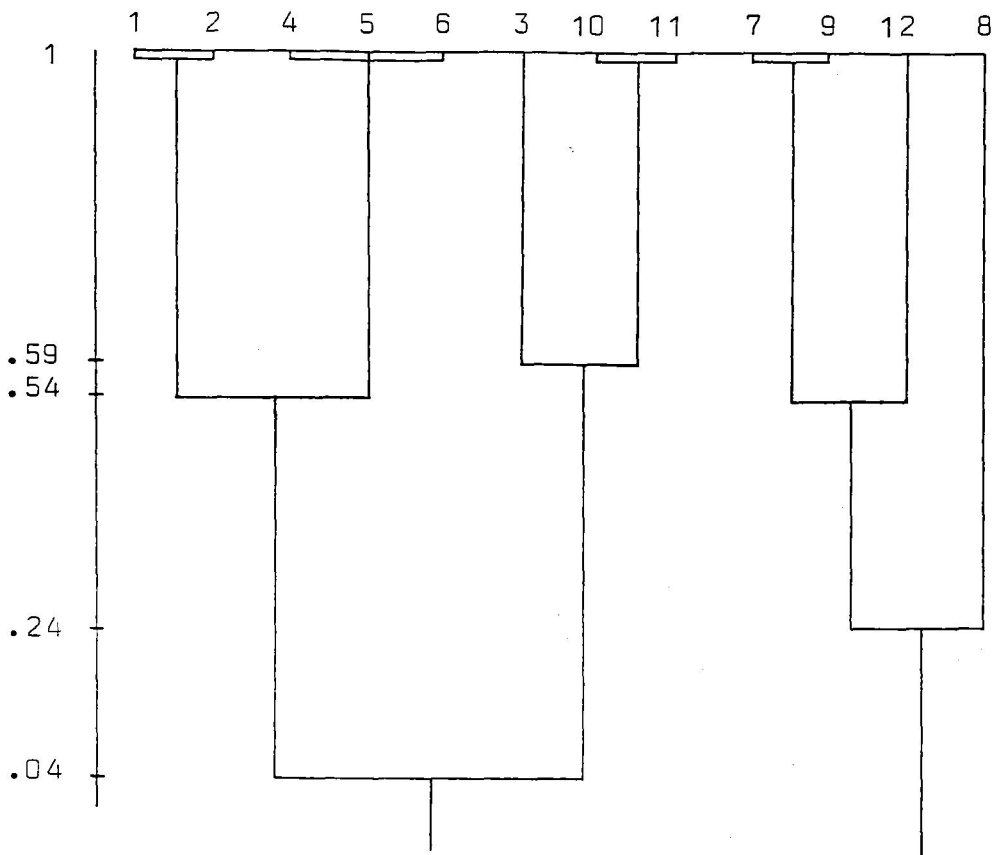
lo de la intersección fuerte.

AGRUPACION. DENDROGRAMAS



Indice S_1 . Tabla 10.

En este caso se alcanza el límite superior 1 ya que existen grupos con todos sus elementos iguales y, por tanto, la intersección fuerte coincide con el campo conjunto de los mismos. La agrupación se muestra bastante correcta, quedando la unidad 8 aislada hasta el final del dendrograma donde se fusiona con los caracteres B. La agrupación de 3 y (10,11) se realiza debido a la escasez relativa del color verde; sin embargo, la agrupación de este grupo con (1,2,4,5,6) formado por caracteres A se realiza con similaridad mínima, es decir, no se agrupan.

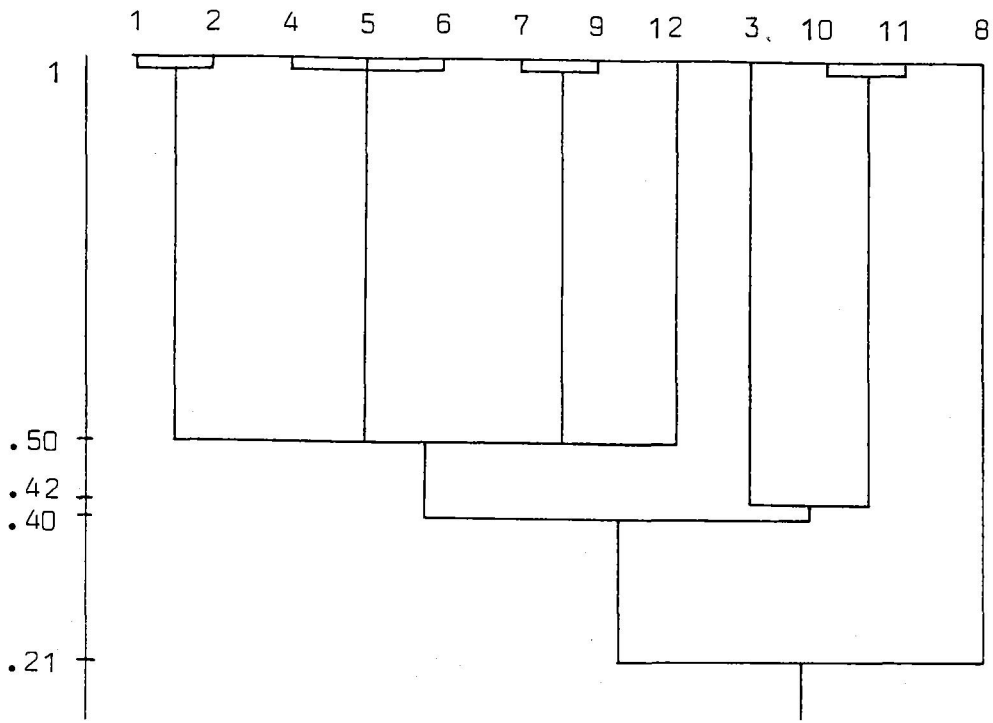


Indice S_1' . Tabla 11.

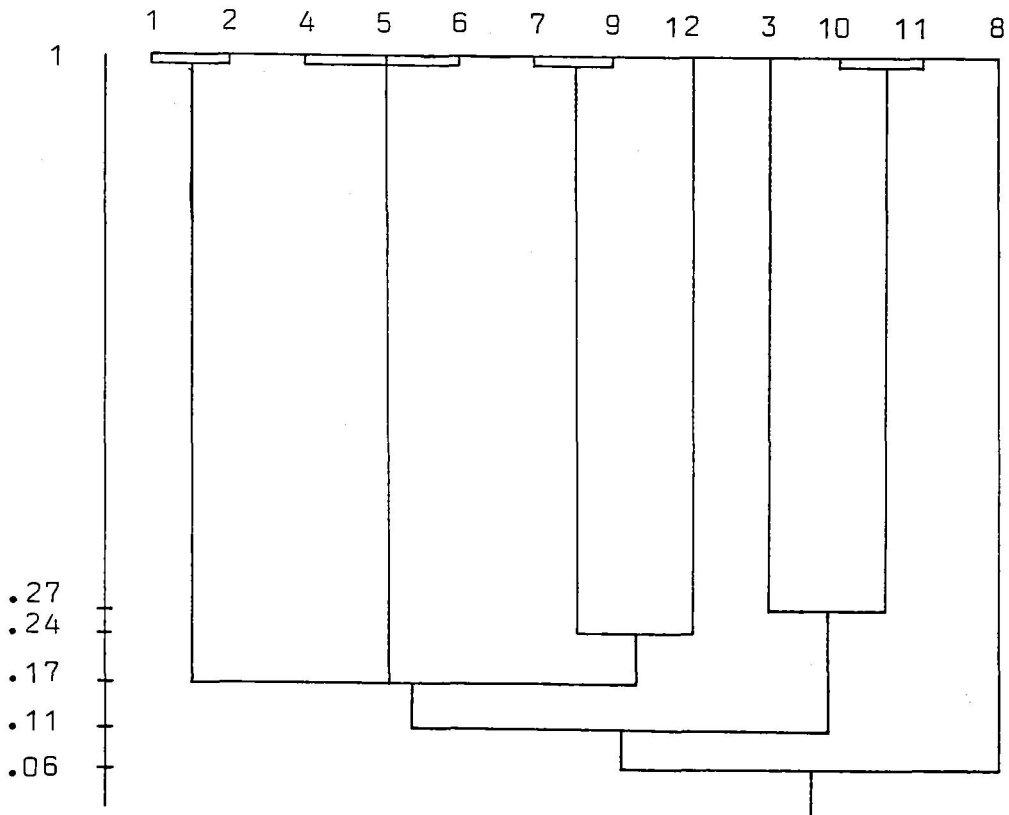
Con ambas medidas S_1 y S_1' se alcanzan, en este caso, los mismos resultados y el campo medio no produce valores extraños.

Considerando la medida S_2 se produce, con bastante intensidad, el fenómeno de que no hay uniformidad en el decrecimiento de la medida de similaridad conforme se desciende en el nivel de la jerarquía. Así se agrupan (1,2) y (4,5,6) con $S_2 = .37$ y (7,9) y 12 con $S_2 = .37$, pero estos grupos se fusionan posteriormente con similaridad $S_2 = .50$. Más adelante ocurre otro caso similar con los grupos (1,2,4,5,6,7,9,12) y (3,10,11), lo que produce problemas en la realización del dendrograma. Sin embargo, estos problemas no aparecen en este caso al considerar las medidas derivadas de S_2 (S_3 y S_4) que toman en cuenta el número de elementos de los grupos que se van a fusionar y los elementos iguales que existen en ellos (este ejemplo abunda, intencionadamente, en dichos casos).

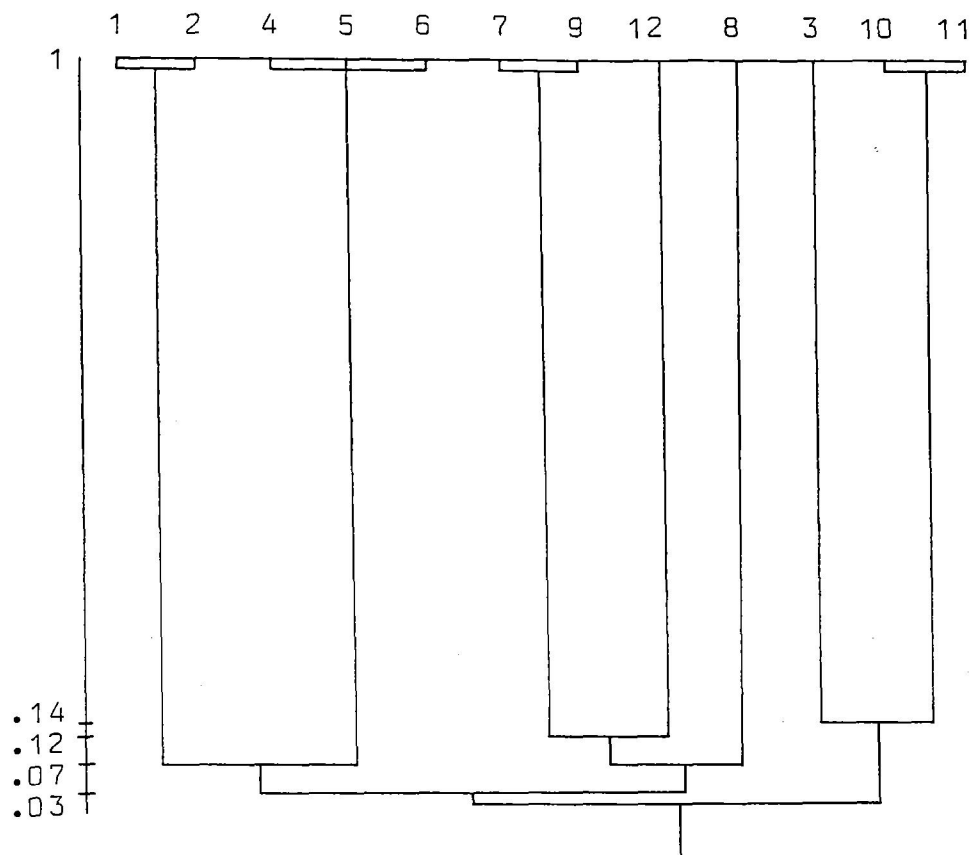
Una aproximación al dendrograma para S_2 es la figura siguiente:



Indice S₂. Tabla 12.



Indice S₃. Tabla 13



Indice S_4 . Tabla 14.

En la tabla 13 (correspondiente a S_3) también aparecen problemas/ debidos al no decrecimiento de la similaridad. Así, previamente se a- / grupan (1,2) y (4,5,6) con $S_3 = .16$ y, en un nivel posterior, se agru- / pan (1,2,4,5,6) y (7,9,12) con $S_3 = .17$. Con el índice S_4 estos incon- / venientes no surgen aunque se observan valores muy bajos del índice / de similaridad debido a la división por el número de elementos del / grupo.

APENDICES

SIMILARIDAD FUERTE

```

10 REM  AGRUPACION DE UNIDADES
20 REM  SIMILARIDAD FUERTE
30 REM  José A. Esquivel Guerrero. 1987 (Enero)
40 CLEAR:max=0:l2=LN(2):CLS
50 CSIZE 2,1:PRINT "          AGRUPACION":CSIZE 0,0
60 INPUT "Número de variables ";v:INPUT "Número de elementos ";p
70 DIM e(v),d(p,v),a(p,p),f(p),fp(p,p):CLS:CSIZE 2,0:PRINT "Número
    ro de estados de las variables ":csize 0,0
80 FOR i=1 TO v
90 READ e(i):PRINT ,"Variable ";i,e(i);" estados":If e(i)> max:
    max=e(i)
100 END FOR i
110 DIM es(v,max),fr(v,max)
120 FOR i=1 TO p
130 READ a$:FOR j=1 TO v:d(i,j)=CODE(a$(j))
140 END FOR i
150 FOR i=1 TO v
160 READ a$:FOR j=1 TO e(i):es(i,j)=CODE(a$(j))
170 END FOR i
180 FOR i=1 TO p
190 FOR j=1 TO v
200 FOR k=1 TO e(j):fr(j,k)=fr(j,k)+1/p*(es(j,k)=d(i,j))
210 END FOR j
220 END FOR i
230 FOR i=1 TO p
240 FOR j=1 TO v
250 FOR k=1 to e(j)
260 IF d(i,j)=es(j,k):f(i)=f(i)-LN(fr(j,k))/e(j):EXIT k
270 END FOR k
280 END FOR j
290 f(i)=f(i)/l2
300 END FOR i
310 REM  Cálculo de atracción

```



```

320 DEFine PROCedure calgrup
330 FOR i=1 TO p-1
340 IF l(i)=i THEN
350 FOR j=i+1 TO p
360 IF l(j)=j:simi i,j
370 END FOR j
380 END IF
390 END FOR i
400 END DEFine calcgrup
410 DEFine PROCedure simi(x,y)
420 DIM d1(v),g(p)
430 in=0:di=0:fa=0
440 FOR u=1 TO p
450 IF l(u)=x OR l(u)=y:in=in+1:g(in)=u
460 END FOR u
470 FOR z=1 TO v
480 FOR k=1 TO e(z)
490 c=0
500 FOR u=1 TO in
510 IF d(g(u),z)=es(z,k) THEN
520 afin
530 FOR u1=1 TO in
540 IF l(g(u))<>l(g(u1)) AND d(g(u1),z)=es(z,k):c=1:EXIT u1
550 END FOR u1
560 EXIT u
570 END IF
580 END FOR u
590 IF c:d1(z)=d1(z)-LN(fr(z,k))/e(z)
600 END FOR k
610 di=di+d1(z)
620 END FOR z
630 di=di/12
640 f1=f(x)+f(y)-di:fp(x,y)=f1:a(x,y)=af/12
650 nu=n(x)+n(y):
660 a(x,y)=a(x,y)/f1
670 END DEFine simi
680 DIM l(p),n(p)

```

```

690 FOR i=1 TO p:l(i)=i:n(i)=1
700 ni=0
710 calgrup
720 REPEAT loop
730 min=0:x1=0
740 FOR i=1 TO p-1
750 IF l(i)=i THEN
760 FOR j=i+1 TO p
770 IF l(j)=j AND a(i,j)>0 AND a(i,j)>min:min=a(i,j):x1=i:y1=j
780 END FOR j
790 END IF
800 END FOR i
810 IF x1=0 OR ni=p-1:PRINT "FIN DE AGRUPACION":STOP
820 PRINT "Nivel ";ni,"Se agrupan ";x1;" y ";y1,"Atracción = ";min
830 recal:esc:ni=ni+1
840 END REPEAT loop
850 DEFINE PROCEDURE recal
860 n(x1)=n(x1)+n(y1):f(x1)=fp(x1,y1)
870 FOR u0=1 TO p
880 IF l(u0)=y1:l(u0)=x1
890 END FOR u0
900 FOR u0=1 TO p
910 IF l(u0)=u0 AND l(u0)<>x1 THEN
920 IF u0<x1:simi u0,x1:ELSE:simi x1,u0
930 END IF
940 END FOR u0
950 END DEFINE recal
960 DEFINE PROCEDURE esc
970 FOR t=1 TO p
980 IF l(t)=t:PRINT INT(f(t)*100+.5)/100;" ";
990 END FOR t
1000 PRINT
1010 FOR t=1 TO p
1020 IF l(t)=t THEN
1030 PRINT "Grupo ";t;" ";
1040 FOR m=t TO p
1050 IF l(m)=t THEN PRINT m;" ";

```

```

1060 END FOR m
1070 PRINT
1080 END IF
1090 END FOR t
1100 FOR k=1 to 1000:END FOR k
1110 END DEFine esc
1120 DEFine PROCEDURE afin
1130 c2=0
1140 FOR el=1 TO in-1
1150 IF d(g(el),z)=d(g(el+1),z) THEN c2=c2+1:ELSE:EXIT el
1160 END FOR el
1170 IF c2=in-1:af=af-LN(fr(z,k))/e(z)
1180 END DEFine afin

```

SIMILARIDAD - AFINIDAD

Basta eliminar la línea 520 completa y, en la línea 640, la expresión $a(x,y)=af/12$.

Además, hay que añadir las líneas:

```
645 a(x,y)=di.
```

MEDIDAS DERIVADAS DE S_2

Hay que añadir la línea 655 IF nrep=1 OR nrep=0 THEN le=1:ELSE: le=LN(nrep)/12 si se trata de S_3 . Para S_4 sirve esta misma línea cambiando la última expresión por $le=nrep$.

También hay que añadir la línea:

```
647 numel
```

y los procedimientos:

```

1200 DEFine PROCedure numel
1210 LOCAl ne(in)
1220 nrep=1
1230 u=1
1240 FOR h=u+1 TO in
1250 IF igual(u,h) THEN nrep=nrep+1
1260 END FOR h
1270 IF nrep<in:nrep=0
1280 END DEFine numel
1290 DEFine FUNction igual(u,h)
1300 FOR m=1 TO v
1310 IF d(g(u),m)<>d(g(h),m):RETurn 0
1320 END FOR m
1330 RETurn 1
1340 END DEFine igual

```

El software se ha realizado en lenguaje estructurado, esto es, sin referencias a bifurcaciones ni subrutinas. El lenguaje admite procedimientos recursivos ó no con ó sin parámetros y llamada a funciones / con iguales características.

Las características del ordenador utilizado son:

- procesador 68000.
- RAM 640K.
- unidad de disquettes 3 1/2 de 720K.

Para que pueda ser utilizado en distintos computadores, la lectura de datos se realiza mediante líneas READ - DATA y la salida es la convencional de pantalla. En cada caso particular habrá que añadir líneas DATA conteniendo el número de estados de cada variable, los estados que definen a cada elemento y los estados de cada variable. En el caso del ejemplo 2 estas líneas serían:

```

10000 DATA 2,2,3
10010 DATA "MAR","MAR","MAV","MAZ","MAZ","MAZ","MBR","mBR","MBR", /
      "MBV","MBV","MBZ"
10020 DATA "Mm","AB","RVZ"

```

CONCLUSIONES

CONCLUSIONES

En este trabajo se han desarrollado nuevas medidas de similaridad definidas en un espacio de patrones establecidos mediante variables / cualitativas multiestado, estudiando sus propiedades y uso en un análisis cluster mediante la construcción de algoritmos de agrupación / que lo llevan a cabo.

La información que aporta una variable se define mediante una nueva función de entropía que toma en cuenta la frecuencia de los estados (como en la entropía clásica) y el número de estados de cada variable (posibilidad sugerida por la forma general de la solución de la ecuación que determina la forma de la entropía). Esta noción se extiende a la información (denominada campo y considerada como perturbación, distorsión, etc.) de unidades y grupos de unidades, con el desarrollo de Teoremas y Corolarios de cálculo y pormenorización de las consecuencias derivadas de variar el número de estados y/o sus frecuencias.

También se ha definido la atracción elemento-grupo y grupo-grupo/ que considera todas las coincidencias entre elementos cualesquiera, y una intersección fuerte que resume la información común a todos y / cada uno de los elementos. El desarrollo de Teoremas y Corolarios permiten estudiar sus propiedades y cotas, realizar los cálculos necesarios y comprobar la verificación de las propiedades de Backer y Jain/ (9) asumidas por Pal y Majumder (4). De estas nociones surgen las nuevas medidas de similaridad relativas al campo de los grupos y su adecuación como medidas de clusterización. Otras medidas derivadas surgen de forma natural, inspiradas tanto en medidas conjuntistas como / en medidas fuzzy.

Estos índices de similaridad se han utilizado para construir un algoritmo global y varios específicos que realizan un análisis cluster jerárquico construido en función de estas nociones. Aplicado este método a la agrupación de un conjunto standard de datos (Michalski y/ Stepp (13)) de microordenadores, se obtienen resultados muy satisfactorios, congruentes con el análisis de estos mismos datos que, median

técnicas de Taxonomía Numérica, aparecen en la literatura al respecto (ver también Dale (20)).

Estas nuevas medidas de similaridad pueden utilizarse en otras aplicaciones de reconocimiento de patrones tales como entropía de un / conjunto fuzzy, fuzzyness, reconocimiento de documentos, clasifica- / ción, etc., y en general donde sea necesario trabajar con índices de / similaridad noparamétricos.

Por construcción, este trabajo se limita a variables cualitativas multiestado pero puede extenderse en otras direcciones que compren- / den:

- extensión a variables ordinales, continuas y definidas por intervalos.
- asignación de pesos a los estados de acuerdo a algún criterio: experiencia del investigador, etc.
- búsqueda de nuevas medidas basadas en la distorsión.
- ampliación a patrones jerárquicamente dependientes mediante un árbol jerárquico.
- ampliación a variables no-independientes.
- árboles de clasificación.

Con respecto a la clusterización, el algoritmo de agrupación se / detiene al llegar a un único grupo, similaridad 0 ó alcanzar un número predeterminado de grupos, pero existen cuestiones que pueden incorporarse, en base a estas ideas, como son: caracterización y estructura de la partición, clasificación de patrones muy disgregados (eliminando en 1ª instancia los más disgregados para reagruparlos posteriormente), optimización de la partición (reclasificación) y validez de / la agrupación, ya sea mediante coeficientes de entropía ó cualquier otro método.

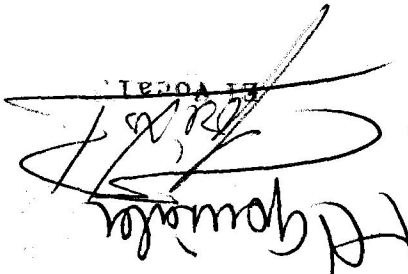
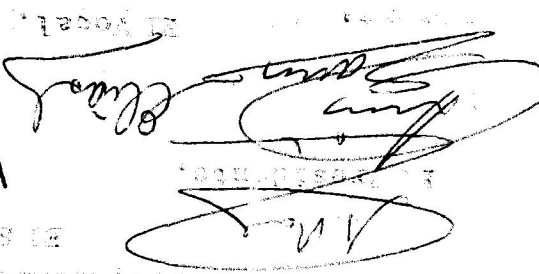
BIBLIOGRAFIA

BIBLIOGRAFIA

- 1) F. M. Reza, An Introduction to Information Theory, Mc Graw-Hill, / New York, 1961.
- 2) D. Dubois and H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York, 1980.
- 3) Q. R. Wang and C. Y. Suen, ' Analysis and Design of a Decision / Tree Based on Entropy Reduction and Its Application to Large Character Set Recognition ', IEEE Trans. Pattern Anal. and Machine / Intell., Vol. PAMI-6, n. 4, pp. 406-417, 1984.
- 4) S. K. Pal and D. Dutta Majumder, Fuzzy Mathematical Approach to / Pattern Recognition, Wiley Eastern, New Delhi, 1985.
- 5) T. Ito, Y. Kodama and J. Toyoda, ' A Similarity Measure Between / Patterns with Nonindependent Attributes ', IEEE Trans. Pattern Anal. and Machine Intell., Vol. PAMI-6, n. 1, pp. 111-115, 1984.
- 6) S. K. Pal and D. Dutta Majumder, ' Fuzzy Sets and Decision-making Approaches in Vowel and Speaker Recognition ', IEEE Trans. Syst., Man and Cyberns., Vol. SMC-7, pp. 625-629, 1977.
- 7) W. X. Xie and S. D. Bedrosian, ' An Information Measure for Fuzzy Sets ', IEEE Trans. Syst., Man and Cyberns., Vol. SMC-14, n. 1, / pp. 151-156, 1984.
- 8) S. K. Pal and B. Chakraborty, ' Fuzzy Set Theoretic Measure for / Automatic Feature Evaluation ', IEEE Trans. Syst., Man and Cybern Vol. SMC-16, n. 5, pp. 754-760, 1986.
- 9) E. Backer and A. K. Jain, ' A Clustering Performance Measure Based on Fuzzy Set Decomposition ', IEEE Trans. Pattern Anal. and / Machine Intell., Vol. PAMI-3, n. 1, 1981.
- 10) A. De Luca and S. Termini, ' A Definition of nonprobabilistic Entropy in the Setting of Fuzzy Set Theory ', Inform. and Control, / Vol. 20, pp. 301-312, 1972.
- 11) M. G. Kendall, Multivariate Analysis, Charles Griffin, London, / 1975.

- 12) E. H. Ruspini, ' A New Approach to Clustering ', Inform and Con- / trol, Vol. 15, pp. 22-32, 1969.
- 13) R. S. Michalski and R. E. Stepp, ' Automated Construction of Clas- / sifications: Conceptual Clustering Versus Numerical Taxonomy ', / IEEE Trans. Pattern Anal. and Machine Intell., Vol. PAMI-5, n. 4, / pp. 396-410, 1983.
- 14) T. Bailey and J. Cowles, ' Cluster Definition by the Optimization / of Simple Measures ', IEEE Trans. Pattern Anal. and Machine In- / tell., Vol. PAMI-6, n. 5, 1984.
- 15) P. H. A. Sneath and R. R. Sokal, Numerical Taxonomy: The Princi- / ples and Practice of Numerical Classification, W. H. Freeman, San / Francisco, 1973.
- 16) L. F. Escudero, Reconocimiento de Patrones, Paraninfo, Madrid, / 1977.
- 17) M. Ben-Bassat and L. Zaidenberg, ' Contextual Template Matching: / A Distance Measure for Patterns with Hierarchically Dependent Fea- / tures ', IEEE Trans. Pattern Anal. and Machine Intell., Vol. / PAMI-6, n.2, 1984.
- 18) E. Diday and J. C. Simon, Clustering Analysis: Communication and / Cybernetics, Vol. 10, Springer Verlag, New York, 1976.
- 19) R. O. Duda and P. E. Hart, Pattern Classification and Scene Analy- / sis, John Wiley, New York, 1973.
- 20) M. B. Dale, ' On the Comparison of Conceptual Clustering and Nume- / rical Taxonomy ', IEEE Trans. Pattern Anal. and Machine Intell., / Vol. PAMI-7, n. 2, 1985.
- 21) S. K. Pal, ' A note on the Quantitative Measure of Image Enhance- / ment Through Fuzzyness ', IEEE Trans. Pattern Anal. and Machine / Intell., Vol. PAMI-4, n. 2, 1982.
- 22) Tao Gu and B. Duibisson, ' A Loose-Pattern Process Approach to / Clustering Fuzzy Data Sets ', IEEE Trans. Pattern Anal. and Machi- / ne Intell., Vol. PAMI-7, n. 3, pp., 66-372, 1985.

- 23) S. Kullback and J. C. Keegel, ' The Minimum Discrimination information Approach to the Analysis of Contingency Tables ', Computational Statistics and Data Analysis, Vol. 3, pp. 133-141, 1985.
- 24) J. E. Doran and F. R. Hodson, Mathematics and Computers in Archaeology, Edinburgh University Press, Edinburgh, 1975.
- 25) T. L. McAuliffe and A. A. Afifi, ' Comparison of a nearest neighbor and other approaches to the detection of space-time clustering ', Computational Statistics and Data Analysis, Vol. 2, / pp. 125-142, 1984.

y Marzo 4 de 1988

D. Antonio Pascual Acosta
 D. Luis Ramon Guayosa
 D. Elias Moreno Bas
 D. José Luis Rino Mejias
 D. Andres Guaiter Casanova

Dpto. con su Raudé
 José Antonio Esquivel Guerrero
 Universidad

Heavido el Tribunal Contador en el día de
 la fecha, constado por:

DILIGENCIA: