# UNIVERSIDAD COMPLUTENSE DE MADRID

## FACULTAD DE CIENCIAS MATEMÁTICAS

**Departamento de Estadistica e Investigacion Operativa**


TESIS DOCTORAL

# Cuestiones notables sobre discriminación de variables cualitativas

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR


**José Miguel García-Santesmases Martín-Tesorero**


**Madrid, 2015**

José Miguel García-Santesmases Martín-Tesorero



# CUESTIONES NOTABLES SOBRE DISCRIMINACION EN VARIBLES CUALITATIVAS

Departamento de Estadística e Investigación Operativa
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
1983

Colección Tesis Doctorales. Nº 255/83


Edita e imprime la Editorial de la Universidad
Complutense de Madrid. Servicio de Reprografía
Noviciado, 3 Madrid-8
Madrid, 1983
Xerox 9200 XB 480
Depósito Legal: M-39683-1983

U N I V E R S I D A D   C O M P L U T E N S E   DE M A D R I D

FACULTAD DE CIENCIAS MATEMATICAS

DEPARTAMENTO DE ESTADISTICA E INVESTIGACION OPERATIVA

# "CUESTIONES NOTABLES SOBRE DISCRIMINACION EN VARIABLES CUALITATIVAS"

Memoria presentada por D. José Miguel García-Santesmases Martín-Tesorero, para optar al grado de Doctor en Ciencias Matemáticas, realizada bajo la dirección del Dr. D. Miguel Sánchez García.

Madrid, Junio, 1982

Quiero expresar mi agradecimiento al profesor Dr. D. Miguel Sánchez García, director de este trabajo, por su consejo y ayuda en todo momento, aún cuando la geografía nos separase y no solo a lo largo de la realización del presente trabajo, si no durante el periodo de mi formación de la cual él ha sido un pilar básico.

También quiero agradecer a mis compañeros del Departamento de Estadística Matemática e Investigación Operativa y a su director, profesor Dr. D. Sixto Ríos García, así como a mis compañeros del Centro de Cálculo de la Universidad Complutense y a todos los que con su interés hicieron posible esta labor. En particular a los profesores Dr. D. Francisco Cano Sevilla y Dr. D. Javier Martín Rodrígo, pues gracias a su estímulo este trabajo ha podido ver su fin.

Por último, quiero agradecer a doña Soledad Estévez por su esmerada mecanografía, que ha hecho posible presentar esta memoria tal y como aquí se hace.

PROLOGO

La presente memoria recoge el desarrollo de una metodología para la discriminación de 2 o más grupos entre sí, cuando estos están descritos en función de variables cualitativas o categóricas.

Si tuvieramos que clasificar en el conjunto de técnicas de tratamiento de datos a esta metodología, la asignaríamos al conjunto de métodos que algunos autores han convenido a llamar Análisis de datos en sentido estricto y que comprendería aquellas técnicas estadísticas que no se apoyan en modelos probabilísticos; frente al conjunto de técnicas ó métodos de la estadística clásica ó técnicas probabilísticas. Esta clasificación no es totalmente excluyente, pudiendo existir métodos de tratamiento de datos que compartan ambas concepciones por lo que sería preferible englobarlas bajo el nombre común de análisis de datos en sentido amplio ó tratamiento de datos.

La presente memoria consta de una introducción y tres capítulos.

En la introducción se hace un breve resumen de los resultados alcanzados, en este campo, fundamentalmente, metodologías basadas en la segunda clase de métodos, poniendo de relieve las dificultades que presentan, sobre todo en su aplicación.

En el primer capítulo se trata de la caracterización, de la solución a un problema de discriminación y se introducen los elementos que van a permitir definir la solución óptima, entre ellos el concepto de bondad de una solución. Se estudian las funciones bondad, que llamamos Bl o funciones bondad compatibles con la relacción de orden parcial definida en el conjunto de soluciones, y se da una condición suficiente para la existencia de estas.

En el segundo capítulo, se trata el problema de la capacidad descriptiva de las soluciones. Se introducen los conceptos de hipótesis y ambigüedad de una hipótesis que servirán para explicitar dicha capacidad descriptiva. Se estudian las relaciones entre las funciones bondad y ambigüedad, caracterizando las bondades basadas en la ambigüedad. Por último, nos centramos en el problema de la caracterización de las funciones bondad que sean de tipo B1 y que estén basadas en la ambigüedad.

En el tercer capítulo se trata el problema del cálculo de la solución. Se introducen los conceptos de segmento cartesiano y segmento terminal, que nos permiten caracterizar máximas locales, para funciones bondad sobre el conjunto de soluciones cartesianas. Se desarrollan tres algoritmos, basados en la idea de segmentación, para la localización de un máximo local. En la segunda parte de este capítulo, se generaliza la definición dada de solución a un problema de discriminación, generalización que surge de forma natural del concepto de segmento terminal, y se propone un algoritmo para calcular una solución generalizada que sea máximo local.

# INDICE

# INTRODUCCION

El desarrollo creciente que en estos últimos años ha experimentado el análisis de datos cualitativos se debe fundamentalmente al desarrollo experimentado a su vez por los ordenadores. Estos han influido en el estadístico en dos aspectos distintos y confluyentes. En primer lugar la potencia de cálculo y de manejo de grandes masas de información que tienen los actuales ordenadores han permitido, a investigadores provenientes en general de campos de las ciencias sociales, biológicas ó del comportamiento tratar situaciones que antes escapaban a sus posibilidades. Muchas de estas situaciones estaban descritas en términos de variables cualitativas y requerían metodologías en el tratamiento que se amoldasen a dichas características y que en general no respondían a los resultados clásicos de la estadística. Esta incapacidad de la estadística de contestar a preguntas concretas planteadas en dichas situaciones llevaba consigo a tratamientos 'pobres' limitados casi exclusivamente a recoger los aspectos descriptivos del problema o lo que era peor a la aplicación de modelos más complicados tipo predicción etc., a situaciones para las que inicialmente no fueron concebidos. La abundancia de estos problemas y de los pobres tratamientos que se les daban fue una de las causas que motivaron el interés por parte de los estadísticos en el estudio de este tema.

Los ordenadores, no solo sirvieron de iman a los problemas y por tanto a sensibilizar a los estadísticos en ellos, sino que a su vez, posibilitaron a los estadísticos su resolución. El ordenador empezó a estar presente en la generación de las nuevas metodologías que requerían dichos problemas no ya solamente a la hora de aplicar a situaciones concretas resultados teóricos, si no que empezaron a jugar un papel importante en la misma concepción de las nuevas teorías y en particular del análisis de datos, pudiendo decir que el desarrollo de este es inseparable del desarrollo de los ordenadores. Sobre este tema pueden consultarse las obras citadas en esta monografía.

Las áreas que mayor interés han despertado dentro de los investigadores en el tema han sido el análisis de las tablas de contingencia múltiple y los modelos log-linear, pudiendose consultar sobre el tema las obras de Bishop, Fienberg y Holland (1978), Haberman (1978), Goodmnan (1978). El análisis de conglomerados, Hartigan (1975), Anderberg (1973), Benzecri (1973), Jambu (1978), Diday (1982) y el análisis de Correspondencias y Correspondencias Múltiples, Benzecri (1973), Lebart (1977).

El tema que estudiamos en el presente trabajo, el análisis discriminante en variables cualitativas, no ha recibido hasta la fecha tanta atención como los temas anteriormente señalados, no existiendo monografías dedicadas exclusivamente al tema, exceptuando la de Goldstein M., Dillon W.R. (1978). Sin embargo, y dada la importancia del tema, sobre todo desde el punto de vista de sus aplicaciones son cada día más numerosas las publicaciones que aparecen sobre él, y debido a esta razón dichas publicaciones no se concentran en una única área sino que aparecen desperdigadas en publicaciones estadísticas, biológicas, investigación de mercados, ingeniería, cibernética, inteligencia artificial, etc.

Tampoco el nombre con que nos referimos al tema del trabajo, discriminación en variables cualitativas es el único nombre que hace referencia a dicho tema. Alguno de los nombres con los que aparece en la literatura son Diagnosis, Reconocimiento de formas, clasificación, asignación, segmentaciónn, ver Gupta R.P. (1980).

Todo problema de discriminación en variables cualitativas puede exponerse en los siguientes términos

Sean N individuos repartidos en k grupos, descritos por p variables cualitativas $\xi_1, \xi_2, \ldots, \xi_p$ de $m_1\ m_2 \ldots m_p$ categorías cada una que llamaremos variables explicativas, y una variable $\eta$, que llamaremos variable grupo con k categorías que se desea explicar basandonos en la información proporcionada por las p variables explicativas.

Dos son los aspectos que se contemplan en todo problema de discriminación y en particular en el que tratamos; siguiendo a Romeder (1973) podemos hablar del aspecto descriptivo del problema, que trata de dar respuestas a preguntas del tipo: ¿Con la información recogida se pueden seperar los grupos de individuos?, ¿son todas las variables explicativas igualmente importante a la hora de separar dichas clases? ¿hay alguna variable o conjunto de variables que influya de forma especial en la definición de una clase?, es decir trata de describir las clases a partir de las variables explicativas $\xi_1 \ldots \xi_p$.

El segundo aspecto, que llamaremos de asignación o clasificación trata el tema de la asignación de nuevos individuos ó individuo no etiquetados, de los que se dispone solamente información sobre sus valores en $\xi_1 \dots \xi_k$ deseando agregarles a alguna de las clases. Resolver este problema lleva consigo la construcción de reglas de decisión o asignación que denominaremos clasificadores.

Aunque los dos aspectos estan muy relacionados, no aparecen así en la literatura sobre el tema, versando en su mayor parte casi todas las publicaciones sobre el aspecto decisional del problema. Bien es verdad que el primer aspecto quedaría cubierto dependiendo del poder descriptivo del clasificador generado y este es el caso de alguno de los clasificadores. Sin embargo es de lamentar que muchos de los resultados conseguidos hasta la fecha no aborden el primer aspecto y lo que es peor no tengan en cuenta la capacidad descriptiva del clasificador que proponen. Ver Luchenbruch (1979), Dillon y Goldstein (1978, 1).

Entre los resultados obtenidos hasta la fecha, podemos citar los siguients

En primer lugar se aplicaron los clasificadores lineales, lo que resultaba forzado pues en definitiva lo que se hacía era imponer una codificación numérica a las variables cualitativas. Sobre las desventajas que presentan dichos métodos puede consultarse Dillon y Goldstein (1978,1), Saporta (1979).

La mayoría de los autores a la hora de tratar el tema han supuesto que la información de que se dispone sobre los grupos o clases a discriminar es de tipo probabilístico y generalmente suponen que los datos corresponden a observaciones de las variables

$\xi_1,\ldots,\xi_p$ a través de un muestreo aleatorio simple, entre ellos consideraremos la solución bayesiana.

Si designamos por $f_i(x)$ la función de densidad de $\xi_1 \ldots \xi_p$ en la clase i-ésima, el clasificador bayes para el problema viene dado, supuesto que la función de pérdida es aquella que penaliza con 1 unidad cada vez que hacemos una mala clasificación, a través de las funciones discriminantes $g_i(x) = f_i(x).p_i$ donde $p_i$ es la probabilidad a priori de la clase i-ésima.

El inconveniente que tiene la solución bayesiana bajo estas hipótesis es de tipo práctico dado que generalmente el número de variables explicativas suele ser relativamente elevado lo que provoca que el número de estados del espacio muestral se haga muy grande dando lugar a que en muchos estados del espacio muestral no haya observaciones y se tenga que hacer una asignación al azar.

A la hora de paliar el problema que presenta la solución bayesiana, y restringiendo el campo a situaciones en las que las variables explicativas sean dicotómicas, diversos autores han propuesto métodos basados en representaciones de las densidades a través de menos parámetros de los que necesita el modelo completo, entre estos destacaremos los modelos propuestos por Bahadur (1961), Martin y Bradley (1972) y Ott y Kronmal (1976).

El modelo de Bahadur consiste en truncar la función de densidad conjunta basandose en la representación:

$$g(x) = g(x_1 \ldots x_p) = \prod_{j=1}^{p} \theta_j^{x_j} (1-\theta_j)^{1-x_j} \{1 + \sum_{j<k} \rho_{jk} z_j z_k + \ldots$$

$$\ldots + \rho_{12\ldots p} z_1 \ldots z_p\}$$

donde $\theta_j = E(\xi_j)$; $z_j = (\xi_j - \theta_j) / (\theta_j(1-\theta_j))^{1/2}$

$\rho_{jk} = E(z_j\ z_k)$ ;... $\rho_{12\ldots p} = E(z_1 \ldots z_p)$

así por ejemplo si aproximamos $g(x)$ truncando las interacciones de orden 2 o superior a dos tendremos

$$g(x) \simeq 1 + \sum_{j<k} \rho_{jk}\ z_j z_k$$

este tipo de solución, resuelve el problema de la asignación en estados donde no hay observaciones pero presenta grandes inconvenientes como son la arbitrariedad de la aproximación, dado que en el ejemplo anterior supondría la no existencia de efectos interactivos, hipótesis en principio demasiado fuerte y la nula capacidad descriptiva del clasificador.

Los modelos propuestos por Martín y Bradley y por Ott y Kronmal están basados en representaciones de densidades mediante combinaciones lineales de polinomios ortogonales. Estos modelos así como el de Bahadur tienen la ventaja de permitir aproximaciones de densidades en función de un número menor de parámetros pero adolecen de los mismos inconvenientes más arriba mencionados.

El inconveniente que presentan las aproximaciones anteriores puede evitarse utilizando la metodología de los modelos log-linear para tratar tablas de contingencia múltiples. En primer lugar este método no está restringido al caso de que las variables explicativas sean dicotómicas, además a la hora de considerar si los términos que representan interacciones de un determinado órden son cero no se hace de forma a priori como en los métodos anteriores si no que se uti lizan tests para contrastar la bondad del ajuste, a la hora de construir modelos que se ajusten a los datos y reduzcan el número de parámetros en la estimación de densidades, permitiendo salvar el problema de la asignación en celdillas o estados del espacio muestral para los que no hubiese observaciones. Este método contempla el problema de la discriminación suponiendo que los datos de que disponemos forman una tabla de contingencia múltiple de dimensión $m_1 \times m_2 \times \ldots \times m_p \times k$ generada bajo un modelo de Poisson ó multinomial que corresponde a un problema de $p+1$ variables de las cuales las $p$ primeras son las variables explicativas y la $p+1$ ésima es la variable grupo.

Si $p_{ij\ldots\ell t}$ designa la probabilidad teórica del estado $ij\ldots\ell t$, se cumple:

$$\log p_{ij\ldots\ell t} = \theta^{12\ldots p\ p+1}_{ij\ldots\ell t} + \theta^{2\ldots p\ p+1}_{j\ldots\ell t} + \ldots + \theta^{1\ 2\ldots p}_{ij\ldots\ell} +$$

$$+ \ldots + \theta^{1}_{i} + \ldots + \theta^{p+1}_{t} + \theta_{o}$$

donde:

$$\sum_{i} \ldots \sum_{t} \theta^{1\ 2\ldots p\ p+1}_{ij\ldots\ell t} = 0 \ ; \ldots ; \sum_{t} \theta^{p+1}_{t} = 0$$

Un determinado modelo reducido que represente la matriz de datos se construye igualando a cero un subconjunto de parámetros $\theta$. Se utilizan estimadores de máxima verosimilitud para estimar los parámetro $\theta$ del modelo reducido y se contrasta la bondad del modelo utilizando tests del cociente de probabilidad. En este sentido pueden consultarse las obras de Goldstein y Dillon (1978); Lachenbruch y Goldstein (1979) y Andersen (1980).

En otra línea de trabajo, pero, persiguiendo el mismo fin, es decir, obtener representaciones parsimoniosas de la situación y así salvar el problema de la desproporción, entre el número de estados del espacio muestral sobre el que se basan los clasificadores, y, el número de observaciones de las que se dispone. Hay que mencionar aquellos métodos que disminuyen el número de estados basándose en una eliminación de variables. Estos métodos se conocen como métodos de discriminación por pasos y se basan en la introducción de variables de forma secuencial ó 'hacia adelante' o eliminación de variables de forma secuencial o 'hacia atrás'. Estos métodos, en efecto, consiguen representaciones parsimoniosas de los datos dado que los criterios de inclusión de variables pueden ser tan restrictivos como se desee y además presentan la ventaja de tener un gran poder descriptivo al seleccionar aquellas variables que realmente importen a la hora de diferenciar los grupos. El gran inconveniente que presentan estas soluciones es que generalmente el n° de variables finales que consideran es tan pequeño comparado al n° total de variables explicativas, que se tiene el peligro de no considerar información clave a la hora de caracterizar los grupos. Así por ejemplo el método debido a Lachin (1973) es similar a los de regresión por pasos buscando la inclusión de variables en el estadístico de independencia con la variable grupo,teniendo en cuenta la información proporcionada por las variables ya incluidas. Sobre este método y aquellos debidos a Dillon y Goldstein, Glick, Rcriffa puede consultarse el capítulo 4 de Dillon y Goldstein (1978).

Un aspecto que no tienen en cuenta estos métodos y que proporcionan buenos resultados a la hora de salvar los inconvenientes que presentan es la posibilidad de reducir el n° inicial de estados del espacio muestral en base a una selección de variables y a una reducción del n° de categorías de las variables de forma que entren aquellas categorias que realmente diferencian los grupos y al mismo tiempo permitiendo el juego a más variables. La reducción de categorias se hace en general en base a solapar diferentes categorías de una misma variable de forma que constituyan una nueva categoría siempre que estas den información similar sobre la variable grupo.

Este aspecto ha sido estimado por Messenger y Mandell (1972), Morgan y Messenger (1972), Kass (1975) y Sánchez (1980) su metodología toma diferentes nombres, entre los más corrientes estan los de segmentación, THAID y Tipología.

En esencia todo proceso de segmentación consiste en particionar el conjunto de individuos observados de forma que cada elemento de esta partición describa de forma óptima la variable grupo. Los elementos de esta partición se definen en base a subconjuntos de valores de las variables explicativas y a partir de estos subconjuntos se procede a hacer predicciones sobre la variable grupo.

La gran potencia que tienen estos métodos a la hora de tratar grandes conjuntos de variables explicativas unido al hecho de tener un gran poder descriptivo, los hacen muy recomendables desde un punto de vista práctico y serán una base importante para el desarrollo del presente trabajo.

Este mismo aspecto ha sido tratado por Saporta (1977, 1979, 1981). Donde propone un método de discriminación eminentemente descriptivo y que desarrolla en dos etapas. En la primera etapa hace una selección progresiva de las variables basandose en el coeficiente de asociación de Thchuprow (1), sobre estos realiza un modelo Log linear con el fin de detectar las interacciones existentes entre estas variables, realizando al mismo tiempo un agrupamiento de categorías para cada variable en base a un criterio de comportamiento similar de estos respecto de la variable grupo.

En la segunda etapa y respecto de las variables creadas en la primera etapa realiza un análisis factorial de correspondencias múltiples de forma que sobre los factores extraidos selecciona aquellos que mejor diferencian los grupos (maximizan la varianza intergrupos) y sobre estos últimos procede a la asignación.

Este método, contempla muchas de las exigencias que imponemos a todo buen método de discriminación, tiene gran poder descriptivo, tanto por el proceso utilizado en la primera etapa, una mezcla del ataque por pasos, segmentación y metodología log-linear los tres con gran capacidad descriptiva, como por la técnica utilizada en la segunda etapa eminentemente descriptiva como es el análisis de correspondencias múltiples. Un defecto importante que puede apreciarse en el método es, que dado que el análisis de correspondencias

---

(1) Para una visión amplia de coeficientes de asociación entre variables cualitativas, recomendamos la obra de Reynolds (1977).

múltiples, es un método que no tiene en cuenta las interacciones de orden mayor o igual a 2 entre las variables explicativas, a la hora de proceder a la clasificación basada en los factores si las variables creadas en la primera etapa tienen efectos no lineales sobre la variable grupo estos pueden quedar obviados.

Por último, para terminar con este abanico de las diferentes metodologías utilizadas para resolver el problema de la discriminación en variables cualitativas, estan, los métodos basados en las distancias. Estos métodos, por su esencia, son métodos que tienen un nulo poder descriptivo sobre la situación que traten, teniendo únicamente un sentido decisional o de asignación.

En general estos métodos están basados en distancias entre distribuciones. En principio estaban diseñados para la asignación de un grupo de individuos en bloque a algúna de las clases, como en Matusita (1955), Goldstein y Dillon (1978) proponen un procedimiento de asignación de individuos, uno a uno, utilizando la distancia de Matusita pero restringiendo su aplicación al caso de tener únicamente dos clases.

# CAPITULO I

## 1.1. Caracterización de la Solución

### 1.1.1. Elementos de un problema de discriminación en variables cualitativas. Definición del problema.

Designaremos por $P$ a un problema de discriminación que vendrá definido por la terna $(I,\eta,E)$ donde:

$I = \{i_1,\ldots,i_N\}$ es el conjunto de individuos o items observados.

$\eta$, variable grupo o variable dependiente ó variable criterio $\eta$ es una variable cualitativa que puede tomar $k$ posibles valores,

$E$, conjunto de variables explicativas, $E = \{\xi_1,\ldots,\xi_p\}$, $\forall \xi_i \in E$, $\xi_i$ es una variable categórica con $m_i$ categorías.

Denotaremos por $K\xi$, siendo $\xi$ una variable cualitativa, al conjunto de cateogrías de esta.

Para todo conjunto, $A$ finito, representaremos por $|A|$ su cardinal. Así con la notación más arriba definida, tendremos:

$$|K\xi_i| = m_i; \qquad |K\eta| = k$$

### 1.1.2. Definición de solución de P.

<u>Definición 1.1.2.1</u>.- Diremos que $S$ es una solución de $P$, si $S$ es una aplicación:

$$S : \prod_{i=1}^{p} K\xi_i \longrightarrow \Delta$$

donde

$$\Delta = \{(p_1,\ldots,p_k) \in R^k \ / \ \sum_{i=1}^{k} p_i = 1, \quad 0 \leq p_i \leq 1; \quad i=1,\ldots,k\}$$

Hemos elegido esta caracterización de la solución al problema P, debido al mayor poder descriptivo que tiene frente a otras caracterizaciones, en efecto:

Las 3 caracterizaciones que podemos dar a una solución de P, vienen dadas por:

- Solución de tipo 1:

Tiene sentido cuando $I$ se puede descomponer en:

$$I = I_1 \cup I_2 \ / \ I_1 \cap I_2 = \emptyset; \quad I_1 \neq \emptyset, \quad I_2 \neq \emptyset$$

Los conjuntos $I_1$ e $I_2$ designan respectivamente al conjunto de individuos o items etiquetados y no etiquetados respectivamente y

$\forall i \in I_1$ conocemos el valor de $\xi_j$ en $i, \xi_j(i)$, $\forall j=1,\ldots,p$ y el valor de $\eta$ en $i$, $\eta(i)$

$\forall i \in I_2$ conocemos el valor de $\xi_j$ en $i$, $\xi_j(i)$, $\forall j=1,\ldots,p$ y desconocemos el valor de $\eta$ en $i$, $\eta(i)$.

Definición 1.1.2.2.- Toda aplicación $S_1 : I_2 \longrightarrow K\eta$ es una solución tipo 1.

- Solución de tipo 2:

Definición 1.1.2.3.- Toda aplicación $S_2 : \prod_{i=1}^{p} K\xi_i \longrightarrow K\eta$ es una solución de tipo 2.

- Solución de tipo 3:

Viene dada en la definición 1.1.2.1.

Dado que las soluciones de tipo 1 no hacen referencia explícita en su definición a las variables explicativas, en general estas soluciones no tendran un poder descriptivo grande, si por este entendemos la capacidad que tiene la solución a contestar preguntas sobre el comportamiento de estas variables. Este tipo de caracterización será válido cuando solo nos interese el aspecto decisional del problema.

Las soluciones de tipo 2, responden al aspecto decisional del problema y mejoran el aspecto descriptivo frente a la solución de tipo 1, dado que $S_2$ se genera a partir del conjunto E. Así toda solución de tipo 2 puede contemplarse como una solución de tipo 1 y además provoca en $\Pi K \xi_i$ una partición en k conjuntos que nos describen los subconjuntos de valores de variables explicativas que favorecen una determinada clase.

La solución que hemos elegido corresponde a las soluciones de tercer tipo, estas mejoran el aspecto descriptivo de las de tipo dos pues las particiones a las que dan lugar hacen referencia al grado de favorabilidad de las clases y no solamente ven si son o no favorables a una determinada clase.

## 1.2. Admisibilidad de la solución

### 1.2.1. Función de semejanza $S$ asociada a P.

Supondremos que la información que I proporciona a la hora de explicar las relaciones existentes entre E y $\eta$ viene representada a través de una función de semejanza, S, entre los conjuntos partes de $I_1$, $P(I_1)$ y $K\eta$.

Definición 1.2.1.1.- Una función de semejanza S asociada al problema P es una aplicación

$$S : P(I_1) \times K\eta \longrightarrow R^+ \cup \{0\}$$

que cumple con las siguientes propiedades:

P.1:

Si $\exists\ \ell \in K\eta$ y $\exists\ I' \subset I_1 \ /\ S(I',\ell) = 0$ entonces
$S(I'',\ell) = 0 \quad \forall I'' \subset I'$

P.2:

$\forall \ell \in K\eta \quad S(I_1,\ell) > 0$

El suponer que a todo problema P podemos asociarle una función $S$, no supone una restricción fuerte desde el punto de vista práctico pues es bastante natural pensar que podemos dar una medida de asociación entre un conjunto de items etiquetados y una determinada clase, dado que una forma de generar estas funciones es mediante funciones de asociación entre item y clase.

Sea

$$\psi : I_1 \times K\eta \longrightarrow R^+ \cup \{0\} \qquad [1.2.1]$$

cumpliendo: $\forall \ell \in K\eta \quad \exists \; i \in I_1 \; / \; \psi(i,\ell) > 0.$

$\psi(i,\ell)$ representa el grado de asociación ó semejanza entre el item $i$ y la clase $\ell$, en general la función $\psi$ vendrá definida por:

$$\begin{aligned} \psi(i,\ell) &= 0 \quad \text{si} \quad \eta(i) \neq \ell \\ \psi(i,\ell) &= 1 \quad \text{si} \quad \eta(i) = \ell \end{aligned} \qquad [1.2.2]$$

es decir, la función $\psi$ será la función de pertenencia del individuo a la clase. Esta función representa una forma bastante natural de medir estas asociaciones en la mayoría de los problemas, donde sobre los elementos de $I_1$, items etiquetados, sabemos con seguridad su pertenencia a alguna de las clases.

No siempre la información que disponemos sobre los individuos etiquetados, será tan precisa como la más arriba señalada, pudiendo pensarse en una función $\psi$ que mida la difusidad o ambiguedad de la pertenencia de un individuo a una determinada clase. Estas situaciones se dan generalmente cuando a la información proporcionada por $I_1$, añadimos la información proporcionada por un conjunto de items, inicialmente no controlados, a los que se les ha aplicado un clasificador de tipo 3 y se desea utilizar esta información a la hora de genera la solución.

A partir de una función $\psi$ definida como en [1.2.1] podemos genera una función de pertenencia $S$ asociada a $\psi$ de la siguiente forma:

$$S_\psi(I',\ell) = \sum_{i \in I'} \psi(i,\ell) \qquad [1.2.3]$$

cuando $\psi$ viene definido por [1.2.2]; $S_\psi(I',\ell) = |\{i \in I \;/\; \eta(i) = \ell\}|$ .

$S$, cumple con las propiedades de la definición 1.2.1.1.

P.1.

$$\forall I' \subset I'' \subset I_1 \Longrightarrow S_\psi(I',\ell) \leq S_\psi(I'',\ell) \leq S_\psi(I_1,\ell)$$

$\forall \ell \in K\eta$, en efecto:

$$S_\psi(I_1,\ell) = \sum_{i \in I_1} \psi(i,\ell) = \sum_{i \in I'} \psi(i,\ell) + \sum_{i \in I_1 - I'} \psi(i,\ell) \geq$$

$$\geq \sum_{i \in I'} \psi(i,\ell) = S_\psi(I',\ell) = \sum_{i \in I''} \psi(i,\ell) +$$

$$+ \sum_{i \in I' - I''} \psi(i,\ell) \geq S_\psi(I'',\ell)$$

P.2.

Es consecuencia del hecho de cumplirse la propiedad exigida a la función $\psi$.

Centraremos la búsqueda de la solución al problema P, supuesto que tenemos definida una función de semejanza $S$ asociada a P. A este par lo designaremos por (P,S).

1.2.2. Coherencia en (P,S).

Entre todas las posibles soluciones que cumplen con la definición 1.1.2.1. vamos a limitarnos a aquellas soluciones de P que sean coherentes con la información que tenemos y que viene definida por la función $S$.

Toda S que cumple con la definición 1.1.2.1. provoca en $\Pi K\xi_i$ una partición que designaremos por $(D_1,D_2,\dots,D_r)$ defini da por:

$$D_i = \{(i_1,\dots,i_p) \in \prod_{i=1}^{p} K\xi_i \ / \ S(i_1,\dots,i_n) = cte\} \qquad [1.2.4]$$

y que denominaremos segmento.

Si $\forall D \subset \Pi K\xi_i$ designamos por $\hat{I}_1(D)$ a:

$$\hat{I}_1(D) = \{i \in I_1 \ / \ \xi_1\dots\xi_k(i) \in D\} \qquad [1.2.5]$$

y por $p_D$ a:

$$p_D = S(i_1,\dots,i_n); \quad (i_1\dots i_n) \in D \qquad [1.2.6]$$

luego $\forall D \subset \Pi K\xi_i \ / \ \hat{I}_1(D) \neq \emptyset$ podemos definir una medida de aso ciación con la clase $\ell$ a través de la función $S$ de la siguien te forma:

<u>Definición 1.2.2.1</u>.- $\forall S$ asociada a P que cumple con la definición 1.2.1.1, definimos:

$$a_S : P(\Pi K\xi_i) \times K\eta \longrightarrow R^+ \cup \{0\}$$

$$a_S(D,\ell) = S(\hat{I}_1(D),\ell) \quad \forall D \ / \ \hat{I}_1(D) \neq \emptyset$$

$a_S(D,\ell)$ no está definida para los D t.q. $\hat{I}_1(D) = \emptyset$

La coherencia de una solución de P del tipo 3 con la función $S$ la haremos a través de la función $a_S$ de la siguien te forma:

<u>Definición 1.2.2.2</u>.- Sea S una solución de P que cumple con la definición 1.1.2.1. Sea $(D_1,D_2,\dots,D_r)$ su partición asocia da.

Diremos que S es coherente con $\mathcal{S}$ si se cumple:

$$a_{\mathcal{S}}(D_i,\ell) \leq a_{\mathcal{S}}(D_i,m) \iff (P_{D_i})_\ell \leq (P_{D_i})_m \qquad \ell,m=1,\ldots,k \qquad [1.2.7]$$

Cuando:

$$\hat{I}_1(D_i) \neq \emptyset, \qquad i=1,\ldots,r \qquad [1.2.8]$$

donde por $(P_{D_i})_\ell$ designamos la $\ell$-ésima componente del vector $P_{D_i}$.

A partir de la definición 1.2.2.2 puede pensarse en varias formas de generar S a través de $\mathcal{S}$ de forma que las soluciones sean coherentes con $\mathcal{S}$, la más inmediata es:

Dado que toda solución S que cumple con la definición 1.1.2.1 está caracterizada por una partición de $\Pi K\xi_i$, $(D_1,D_2,\ldots,D_r)$; a la hora de definir una solución de P, lo haremos a través de su partición asociada.

Designamos por S la solución de P que vamos a generar a través de $\mathcal{S}$ si $(D_1,\ldots,D_r)$ es su partición asociada cumpliendo con [1.2.8].

$\forall i$, $i=1,\ldots,r$, construimos el conjunto $AD_i = \{a_{\mathcal{S}}(D_i,1),\ldots$ $\ldots,a_{\mathcal{S}}(D_i,k)\}$.

Sea $r(a_1(D_i,1))$ el rango del elemento $A_{\mathcal{S}}(D_i,1)$ en el conjunto $AD_i$.

Sea $r_{AD_i} = \sum_{\ell=1}^{k} r(a_{\mathcal{S}}(D_i,\ell))$

y la solución S queda definida por:

$$(P_{D_i})_\ell = \frac{r(a_{\mathcal{S}}(D_i,\ell))}{r_{AD_i}} \qquad [1.2.9]$$

Es trivial comprobar que la $S$ así definida cumple con [1.2.7].

Designamos por $C(P,S)$ al conjunto de soluciones de $P$ cumpliendo con la definición 1.1.2.1 y que sean coherente con $S$.

1.2.3. Soluciones admisibles.

Entre todas las soluciones de $C(P,S)$, nos interesa caracterizar aquellas soluciones que además de mantener el orden que la función $a_S$ proporcionaba a los pares $(D_i,\ell)$ conserve también la distancia en esa ordenación. Entenderemos que una solución es admisible cuando conserve dichas distancias; definiendo:

<u>Definición 1.2.3.1</u>.- $\forall S \in C(P,S)$ $S$ es admisible cuando:

$((P_{D_i})_1,(P_{D_i})_2,\dots (P_{D_i})_k)$ es proporcional a

$(a_S(D_i,1),\dots,a_S(D_i,k)) \quad \forall i=1,\dots,r$

Donde $(D_1,\dots,D_r)$ es la partición asociada a $S$.

Al conjunto de soluciones admisibles lo designamos por $A(P,S)$.

1.2.3.1. Propiedades de $A(P,S)$.

P.1. $\forall S \in A(P,S)$, $S$ viene definida por las relaciones:

$$(P_{D_i})_\ell = \frac{a_S(D_i,\ell)}{\sum_{\ell=1}^{k} a_S(D_i,\ell)} \qquad \begin{array}{l} i=1,\dots,r \\ \ell=1,\dots,k \end{array} \qquad [1.2.10]$$

donde $(D_1,\ldots,D_r)$ es la partición asociada a S.

Es consecuencia inmediata de la definición

P.2. $A(P,S) \neq \emptyset$.

Demostración: Consideramos la solución de P cuya partición asociada contiene un único elemento, $\Pi K\xi_i$. Esta solución es admisible para $(P,S)$ $\forall S$ cumpliendo con la definición 1.2.1.1.

En efecto:

$$\hat{I}_1(\prod_{i=1} K\xi_i) = I_1 \neq \emptyset$$

$S(I_1,\ell) > 0$ $\forall \ell \in K\eta$ por la P.2 de la definición 1.2.1.1.

luego $\sum_{\ell=1}^{k} a\ (\Pi K\xi_i,\ell) > 0$ pudiendose aplicar [1.2.10].

1.2.4. Soluciones g-admisibles.

1.2.4.1. Función peso.

Definición 1.2.4.1.- Diremos que f es una función de peso si f es una función:

$$f:\ P(\Pi K\xi_i) \longrightarrow R^+ \cup \{0\}$$

cumpliendo con las propiedades:

P.1 $\forall D_1,D_2 \subset \Pi K\xi_i$ si $|\hat{I}_1(D_1)| \leq |\hat{I}_1(D_2)|$ entonces $f(D_1) \leq (D_2)$

P.2 $f(\Pi K\xi_i) < \infty$

P.3 $f(D) = 0 \iff \hat{I}_1(D) = \emptyset$

P.4 $\forall D_1,D_2 \subset \Pi K\xi_i \ /\ D_1 \cap D_2 = \emptyset \implies f(D_1 \cup D_2) = f(D_1) + f(D_2)$.

Entre estas funciones $f$ podemos considerar como más corrientes:

$$f_1(D) = |\hat{I}_1(D)| \qquad [1.2.11]$$

$$f_2(D) = \frac{|\hat{I}_1(D)|}{|I_1|} \qquad [1.2.12]$$

$$f_3(D) = K|\hat{I}_1(D)|; \quad K \in R^+ \qquad [1.2.13]$$

La introducción de la función de peso en $P(\Pi K\xi_i)$, nos va a servir para considerar la importancia de los segmentos que definen las soluciones como factores a tener en cuenta a la hora de encontrar la solución óptima. Así introducimos el concepto de solución g-admisible como aquella solución admisible cuyos segmentos tienen peso mayor o igual a una cierta cota. Así definimos:

<u>Definición 1.2.4.2</u>.- El conjunto de funciones g-admisibles lo designamos por $A_g(P,S)$ y viene definido por:

$$A_g(P,S) = \{S \in A(P,S) \ / \ f(D_i) \geq g, \ (D_1,\ldots,D_r) \text{ es la partición asociada a } S\} \quad \text{donde } g \in R^+$$

A partir de la definición 1.2.4.2. se deducen las siguientes propiedades:

P.1 $\quad A_{g_2}(P,S) \subset A_{g_1}(P,S) \qquad \forall g_1 \leq g_2$

P.2 $\quad A_o(P,S) = A(P,S)$

P.3 $\quad A_g(P,S) = \emptyset \qquad \forall g > f(\Pi K\xi_i)$

## 1.3. Sobre la solución óptima

Una vez delimitado el conjunto de posibles soluciones al problema de discriminación $(P,S)$, se trata de encontrar entre todas las posibles aquellas solución que sea la mejor. El concepto de mejor solución ó solución óptima dependerá del criterio que fijemos para decidir cuando una solución es mejor que otra. En definitiva se tratará de dotar de una relación de orden al conjunto $A_g(P,S)$ y de encontrar el medio de llegar al máximo. Entre todas las posibles formas de dotar de una relación de orden a $A_g(P,S)$ nos ceñiremos a considerar aquellas que vienen definidas mediante funciones de $A_g(P,S)$ en $R$ y valernos de la relación de orden definida en $R$. Estas funciones las denominaremos funciones bondad.

Definición 1.3.1.- Toda aplicación

$$B : A_g(P,S) \longrightarrow R$$

es una función bondad para $(P,S)$.

La función $B$ provoca en $A_g(P,S)$ una relación de orden definida por:

$\forall S_1,S_2 \in A_g(P,S)$, $S_1$ es preferido o indiferente a $S_2$ si $B(S_1) \geq B(S_2)$.

Definición 1.3.2.- Diremos que $S^* \in A_g(P,S)$ es B-óptima para $(P,S)$ si

$$B(S^*) = \underset{S \in A_g(P,S)}{\text{Máx}} B(S)$$

Otro elemento que nos servirá para caracterizar a la solución óptima, será el cardinal de la partición asociada a esta, y que denominaremos tamaño.

<u>Definición 1.3.3</u>.- $\forall S \in A_g(P,S)$ / $(D_1,\ldots,D_r)$ es su partición asociada, definimos tamaño de $S$, $\tau(S)$:

$$\tau(S) = |\{D_1,\ldots,D_r\}| = r$$

El tamaño de una solución nos servirá para elegir entre las soluciones con la misma bondad aquella que tenga menor tamaño. El tamaño y la bondad de la solución constituyen lo que denominaremos Parsimonia de una solución.

<u>Definición 1.3.4</u>.- $\forall S \in A_g(P,S)$, definimos parsimonia de $S$, $M(S)$

$$M(S) = (B(S),\tau(S))$$

Identificaremos mejor solución o solución óptima como la solución más parsimoniosa.

Varios son los caminos que utilizaremos para encontrar dicha solución, entre estos consideramos:

1) Encontrar el conjunto de soluciones B-óptimas y para este conjunto la solución con tamaño mínimo.

2) Fijar un tamaño máximo y encontrar las B-óptimas para dicho tamaño.

3) Definir una función real de la parsimonia y encontrar la solución que maximice dicha función.

Al desarrollo del primer punto y a la selección de distintas funciones B dedicaremos el resto de la memoria.

## 1.4. Funciones Bondad Bl.

A la hora de establecer criterios que nos generen funciones bondad, uno de los más naturales está basado en el siguiente principio: "cuanto más información tengamos sobre las variables explicativas mediante S, mejor será la capacidad predictiva de S y por tanto la bondad de S". Este principio lo traducimos en términos de funciones bondad, a través de la siguiente relacción de orden definida en $A_g(P,S)$.

Definición 1.4.1.- Para todo $s_1, s_2 \in A_g(P,S)$, diremos que $s_2$ es preferido o indiferente a $s_1$, $s_1 \precsim s_2$, cuando la partición asociada a $s_1$ se deduce de la partición asociada a $s_2$ uniendo dos o más elementos de la $s_2$.

Esta relación de orden es un orden parcial en $A_g(P,S)$.

Bajo este criterio el problema de encontrar la solución óptima estaría resuelto siempre que exista el elemento máximo de $(A_g(P,S), \precsim)$. La siguiente propiedad, de demostración inmediata a partir de la definición 1.4.1 nos ilustra este aspecto:

Propiedad 1.4.1.- Si la partición $(\{i_1,\ldots,i_p\};\ (i_1,\ldots,i_p) \in \Pi K\xi_i)$ define una solución de $A_g(P,S)$ entonces dicha solución es máximo de $(A_g(P,S), \precsim)$ y por tanto define la solución óptima.

En general dicha solución no existirá en el conjunto $A_g(P,S)$ y por tanto el problema de localización del óptimo se tendrá que formular en los términos del apartado 1.3, teniendo que definir funciones bondad que sean compatibles con el orden, $\precsim$, definido en $A_g(P,S)$.

En caso que exista la solución óptima obtenida directamente a partir de $\lesssim$ , no siempre esta será la más parsimoniosa pues puede que su tamaño, que en este caso coincidiría con $|\Pi K\xi_i|$ fuese demasiado grande, interesandonos atacar el problema de obtención del óptimo, fijando un tamaño máximo a las soluciones. En este caso también aparece necesaria la introducción de la función bondad que sea compatible con $\lesssim$ .

A las funciones bondad que sean compatibles con $\lesssim$ las denominaremos funciones Bl.

<u>Definición 1.4.2</u>.- Toda función bondad, B, que cumple con la definición 1.3.1 es Bl cuando se cumple:

$$\forall S_1, S_2 \in A_g(P,S) \;/\; S_1 \lesssim S_2 \Longrightarrow B(S_1) \leq B(S_2)$$

A partir de la definición 1.4.2 es inmediata la siguiente propiedad:

<u>Propiedad 1.4.2</u>.- Si existe el máximo de $(A_g(P,S), \lesssim)$, este coincide con la B-óptima para cualquier B que sea Bl.

1.4.1. Caracterización de las funciones Bl

Centraremos nuestra atención en aquellas situaciones $(P,S)$ donde $S$ viene definida como en [1.2.3] generada a partir de una función $\psi$.

1.4.1.1. Conjunto $M_r$, $M$.

Designaremos por $M_r$ el conjunto de las matrices en $R^+ \cup \{0\}$ de dimensión $r \times k$ tales que no tengan filas ni columnas nulas.

Sea $M = \bigcup_{r=1}^{|\Pi K\xi_i|} M_r$.

$\forall S \in A_g(P,S)$ le asociamos un elemento de $M$ a través de la correspondiencia $\Phi$ definida por:

Definición 1.4.1.1.- $\phi : A_g(P,S) \longrightarrow M$

$$\phi(s) = M_s \in M_r \;/\; r \text{ es el tamaño de } S$$

y $M_s$ viene definida por:

Sea $(D_1,\ldots,D_r)$ la partición asociada a $S$, el elemento $(i,j)$ de $M_s$ viene definido por:

$$m_{ij} = a_S(D_i,j); \qquad \begin{matrix} i=1,\ldots,r \\ j=1,\ldots,k \end{matrix} \qquad [1.4.1]$$

La definición de la correspondencia $\phi$ nos va a permitir caracterizar las funciones bondad Bl a partir de funciones definida en $M$.

La relación de orden $\lesssim$ definida en $A_g(P,S)$, se traduce en $M$ de la siguiente forma:

$\forall S_1,S_2 \in A_g(P,S) \;/\; S_1 \lesssim S_2$ se cumple que $S_1$ se deduce de $S_2$ uniendo elementos de la partición asociada a $S_2$ para formar la partición asociada a $S_1$, debido a [1.4.1], $\phi(S_1)$ se deduce de $\phi(S_2)$ sumando las filas de $\phi(S_2)$ correspondientes a los elementos que se unen de la partición de $S_2$, para formar las filas de la matriz $\phi(S_1)$.

1.4.1.2. Operadores X y X'

En $M_r$ definimos los siguientes operadores

Definición 1.4.1.2.- $\forall \ell \leq r,\quad X_{i_1 i_2 \ldots i_\ell} : \mathcal{M}_r \longrightarrow \mathcal{M}_{r-(\ell-1)}$

$$M \longrightarrow X_{i_1 i_2 \ldots i} \; M$$

Donde $X_{i_1 \ldots i_\ell}\,M$ se deduce de $M$, eliminando las filas $i_1 \ldots i_\ell$ y sustituyendolas por su suma, y dejandola como fila $r-(\ell-1)$-ésima.

El operador $X_{i_1 \ldots i_\ell}$ se designará por $X$ cuando no haya necesidad de hacer referencia explícita a los índices $i_1 \ldots i_\ell$.

Asociado al operador $X_{i_1 \ldots i_\ell}$, definimos el operador $X'_{i_1 \ldots}$

Definición 1.4.1.3.- $\forall X_{i_1 \ldots i_\ell}$ definido de $\mathcal{M}_r$ en $\mathcal{M}_{r-(\ell-1)}$ definimos el operador:

$$X'_{i_1 \ldots i_\ell} : \mathcal{M}_r \longrightarrow \mathcal{M}_\ell$$
$$M \longrightarrow X'_{i_1 \ldots i_\ell}\,M$$

donde $X'_{i_1 \ldots i_\ell}\,M$ está formada por la matriz cuyas filas son las $i_1, \ldots i_\ell$ de $M$.

Análogamente a $X$, designaremos por $X'$, el operador $X'_{i_1 \ldots i_\ell}$ cuando no sea necesario hacer referencia explícita a los í dices que lo definen.

Un caso particular importante de estos operadores y que desig remos por $U$ viene definido por:

$\forall r \leq |\Pi K \xi_i|; \qquad U : \mathcal{M}_r \longrightarrow \mathcal{M}_1$

El único operador posible que se aplica en $\mathcal{M}_1$ es el $X_{1\;2\ldots}$ luego $\forall r \leq |\Pi K \xi_i|; \qquad U = X_{1\;2\;\ldots\;r}$ [1.4.2]

Son inmediatas las siguientes propiedades

P.1 $\forall M \in \mathcal{M}; \quad U'M = M$

P.2 $\forall S_1, S_2 \in A_g(P,S); \quad U\phi(S_1) = U\phi(S_2) = \phi(\{\Pi K \xi_i\})$ [1.4.3]

De las definiciones 1.4.1, 1.4.1.1, 1.4.1.2 es inmediato el siguiente resultado:

$\forall S_1, S_2 \in A_g(P,S)$ / $S_1 \precsim S_2$ $\exists$ una sucesión de operadores $X^1, \ldots, X^h$ t.q. $\phi(S_1) = X^1(X^2(\ldots(X^h\phi(S_2)) \ldots ))$ [1.4.4]

1.4.1.3. Conjunto $\mathcal{W}$

Definimos en $\mathcal{M}$ el conjunto de funciones $\mathcal{W}$ definido por

<u>Definición 1.4.1.4</u>.- $\mathcal{W} = \{W : \mathcal{M} \longrightarrow R^+ \cup \{0\}\}$ que cumplen las propiedades $P_1, P_2, P_3\}$

P.1 $W(XM) \leq W(M) \quad \forall M \in \mathcal{M}$

P.2 $\forall M \in \mathcal{M}$ / si $X'M$ tiene todas sus filas proporcionales entonces $W(XM) = W(M)$

P.3 $\forall M_1, M_2 \in \mathcal{M}$ / $M_1$ es igual a $M_2$ salvo una reordenación de filas, entonces $W(M_1) = W(M_2)$.

<u>Teorema 1.4.1.1</u>.- La condición necesaria y suficiente para que exista B de tipo Bl es la existencia de $W \in \mathcal{W}$.

<u>Demostración</u>: Es condición necesaria:

Sea B función bondad Bl, es fácil comprobar que:

$\forall M \in \mathcal{M}\ \exists (P,S)$ t.q. $\exists\ S \in A_g(P,S)/\phi(S) = M$

entonces, definimos W:

$$W(M) = B(S)$$

y la W así definida es de $\mathcal{W}$.

En efecto veamos que cumple con las propiedades:

P.1 $\forall M \in \mathcal{M}$ y $\forall X$, operador como el definido en 1.4.1.2 se si $S$ es t.q. $\phi(S) = M$ entonces $\exists\ S_1/S_1 \lesssim S$ y $\phi(S_1) = XM$, luego $W(M) = B(S) \geq B(S_1) = W(XM)$

P.2 $\forall M \in \mathcal{M}$ si $\exists\ X\ /\ X'M$ tiene todas sus filas proporcionales, si $S$ es t.q. $\phi(S) = M$, entonces $\phi(S) = XM$, por definición 1.4.1.1 y [1.2.10] y [1.2.4].

Luego $W(M) = B(S) = W(XM)$

P.3 Es evidente a partir de la definición de W.

Es condición suficiente:

Sea $(P,S)$, y consideremos $S \in A_g(P,S)$

Si $\exists\ W \in \mathcal{W}$, $\exists B$ de tipo Bl para $A_g(P,S)$.

$\forall S \in A_g(P,S)$, definimos $B(S) = W(\phi(S))$

La B así definida cumple con ser de Bl.

En efecto, por [1.4.4]

$\forall S_1, S_2\ /\ S_1 \lesssim S_2\ \ \exists\ X^{(1}, \ldots, X^{(h}$ t.q. $\phi(S_1) = X^1(\ldots$

$\ldots(X^h\phi(S_2)\ldots)$

Por ser $W \in \mathcal{W}$ se tiene:

$$B(S_1) = W(\phi(S_1)) = W(X^{(1}(X^{(2}\dots(X^{(h}\phi(S_2))\ \dots)) \leq$$
$$\leq W(X^{(2}\ \dots\ (X^{(h}\phi(S_2))\ \dots) \leq \dots \leq W(X^{(h}\phi(S_2)) \leq$$
$$\leq W(\phi(S_2)) = B(S_2).$$

El resultado anterior, nos será de gran utilidad a la hora de caracterizar las funciones Bl; bastará con caracterizar el conjunto de funciones $\mathcal{W}$.

1.4.1.4. Variabilidad interna de una matriz.

Sea $W \in \mathcal{W}$, definimos la variabilidad interna de $M \in \mathcal{M}$, y la designamos por $V(M)$ a la función:

$$V : \mathcal{M} \longrightarrow R$$
$$V(M) = W(M) - W(UM)$$

donde $U$ es el operador definido en [1.4.2].

Es inmediata la siguiente propiedad:

P.1 $V(M) \geq 0 \quad \forall M \in \mathcal{M}$

Esta definición de la función, V, va a representar la variabilidad interna de la matriz M, y en términos de la solución nos va a representar el aumento experimentado en la bondad de una solución, respecto de la bondad de la solución trivial, que viene definida por $\{\Pi K\xi_i\}$, esto es la solución que tendríamos para (P,S) sin utilizar la información de la variables explicativas.

Nos centraremos en aquellas funciones W de $\mathcal{W}$ cuya V asociada cumpla con la siguiente ecuación:

$\forall M \in \mathcal{M}$, $\forall$ operador $X$.

$$V(M) = V(XM) + V(X'M) \qquad [1.4.5]$$

La justificación de esta restricción impuesta a $\mathcal{W}$, reside en el significado de la anterior igualdad. Esta nos dice que la variabilidad interna de una solución se puede descomponer en la variabilidad de la solución obtenida de la unión de dos o más segmentos de la primera más la variabilidad interna de dichos segmentos.

Designaremos por $\mathcal{W}^*$ el subconjunto de $\mathcal{W}$ cuyos elementos tienen asociados funciones $V$ que cumplen con [1.4.5].

Se cumple el siguiente resultado:

Lema 1.4.1.1.- $\forall W \in \mathcal{W}$, $\forall M \in \mathcal{M}$, $\forall$ operador $X$ [1.4.6].
$W \in \mathcal{W}^* \Longleftrightarrow W(M) = W(XM) + W(X'M) - W(U(X'M))$.

Demostración.- $W \in \mathcal{W}^* \Longleftrightarrow V(M) = W(M) - W(UM) = V(XM) + V(X'M)$, $\forall M \in \mathcal{M}$, $\forall$ operador $X \Longleftrightarrow W(M) = W(XM) - W(U(XM)) + W(X'M) - W(U(X'M)) + W(UM) \Longleftrightarrow W(M) = W(XM) + W(X'M) - W(U(X'M) + (W(UM) - W(U(XM))) \Longleftrightarrow W(M) = W(XM) + W(X'M) - W(U(X'M))$ por [1.4.3], $\forall M \in \mathcal{M}$, operador $X$.

El resultado del lema anterior nos va a permitir caracterizar las funciones de $\mathcal{W}^*$ en función de los valores que tomen $\mathcal{M}_1$ y $\mathcal{M}_2$.

En efecto:

Lema 1.4.1.2.- $\forall W \in \mathcal{W}^*$, $W$ queda definida conociendo sus valores en $\mathcal{M}_1$ y $\mathcal{M}_2$.

Demostración.- Supongamos que $W$ definida en $\mathcal{M}$ que cumplen con las propiedades que caracterizan $\mathcal{W}^*$, veamos que $\forall M \in \mathcal{M}_r$ $W(M)$ aparece en función de elementos de $\mathcal{M}_1$ y $\mathcal{M}_2$.

1) Para $r = 3$, $M \in \mathcal{M}_3$

Construimos: $M_1 = X_{12}M \in \mathcal{M}_2$, $X'_{12}M \in \mathcal{M}_2$ .

y se cumple: $W(M) = W(X_{12}M) + W(X'_{12}M) - W(U(X'_{12}M))$

$$= W(M_1) + W(X'_{12}M) - W(U(X'_{12}M)).$$

2) Supuesto que se cumple para $r = \ell$ lo veremos para $\ell+1$

Sea $M \in \mathcal{M}_{\ell+1}$

entonces:

$$X_{12}M \in \mathcal{M}_\ell$$

$$X'_{12}M \in \mathcal{M}_2$$

Luego estan definidas $W(X_{12}M)$, $W(X'_{12}M)$, $W(M(X'_{12}M))$ en función de elementos de $\mathcal{M}_1$ y $\mathcal{M}_2$.

Y se cumple:

$$W(M) = W(X_{12}M) + W(X'_{12}M) - W(U(X'_{12}M))$$

De 1) y 2) podemos deducir que $W(M)$ puede ponerse en función de elementos de $\mathcal{M}_1$ y $\mathcal{M}_2$.

El resultado del lema anterior, lo hemos utilizado para preguntarnos, ¿cuáles son las condiciones que tendrá que cumplir una función definida en $\mathcal{M}_1 \cup \mathcal{M}_2$ para que pueda genetar una función $W$ de $\mathcal{W}^*$?.

La contestación a esta pregunta la damos en el lema 1.4.1.3, donde la caracterización de $W^*$ se hace a través de funciones aditivas, sugeridas a la vista de los resultado del lema 1.4.1.2.

Lema 1.4.1.3.- Sea $\mathcal{V} = \{v : R^{+k} - \{0\} \longrightarrow R^{+} \cup \{0\} \,/\, v(x+y) \leq v(x) + v(y),\ v(\lambda x) = \lambda v(x)\ \forall \lambda \in R^{+}\}$

Entonces $\exists$ una aplicación $\gamma$:

$$\gamma : \mathcal{V} \longrightarrow \mathcal{W}^{*}$$

Demostración.- $\forall v \in \mathcal{V}$ definimos $\gamma(v) = W$ donde:

$$\forall M \in \mathcal{M}_1; \qquad W(M) = v(M)$$

$\forall M \in \mathcal{M}_r$, designamos por $m_1, m_2, \ldots, m_r$ las $r$ filas de $M$ y definimos:

$$W(M) = v(m_1) + v(m_2) + \ldots + v(m_r) \qquad [1.4.7]$$

La $W$ así definida es de $\mathcal{W}^{*}$, dado que cumple con las propiedades de $\mathcal{W}^{*}$, como vemos a continuación:

P.3 Si $M_1, M_2 \in \mathcal{M}$ son iguales salvo una reordenación de sus filas. Entonces $W(M_1) = W(M_2)$, es inmediato debido a [1.4.7].

P.1 $W(XM) \leq W(M)$ $\forall$ $M \in \mathcal{M}$, $\forall$ operador $X$.

Será suficiente con demostrarlos para $X_{12}$

$$W(X_{12}M) = v(m_1+m_2) + v(m_3) + \ldots + v(m_r) \leq \sum_{i=1}^{r} v(m_i) = W(M)$$

P.2 $\forall M \in \mathcal{M}$ si $\exists$ un operador $X$ / $X'M$ tiene sus filas proporcionadas, $W(XM) = W(M)$

Supongamos sin pérdida de generalidad que $M \in \mathcal{M}_r$ y las $\ell$ primeras filas son todas proporcionales, entonces $\exists\ \lambda_1, \lambda_2, \ldots, \lambda_\ell \in R^{+}$ t.q. $m_i = \lambda_i m$ $\forall i = 1, \ldots, \ell$.

$$W(X_{12\ldots\ell}M) = v(m_1+\ldots+m_\ell)+v(m_{\ell+1}) +\ldots+ v(m_r) =$$

$$= v(\sum_{i=1}^{\ell} \lambda_i m) + v(m_{\ell+1}) +\ldots+ v(m_r) =$$

$$= v(m.\Sigma\lambda_i) + v(m_{\ell+1}) +\ldots+ v(m_r) =$$

$$= \Sigma\lambda_i v(m) + v(m_{\ell+1}) +\ldots+ v(m_r) = \sum_{i=1}^{r} v(m_i) = W(M)$$

Luego la $W$ definida por [1.4.7] es de $\mathcal{W}$, veamos que cumple con la ecuación [1.4.6]

$\forall M \in \mathcal{M}$, $\forall$ operador $X$ de $M$, que supondremos $X_{i_1\ldots i_\ell}$ y $M \in \mathcal{M}_r$.

$$W(X_{i_1\ldots i_\ell}M) = \sum_{\substack{i=1 \\ i\notin\{i_1,\ldots,i_\ell\}}}^{r} v(m_i) + v\left(\sum_{j=i_1}^{i_\ell} m_j\right)$$

$$W(X'_{i_1\ldots i_\ell}M) = \sum_{j=i_1}^{i_\ell} v(m_j)$$

$$W(U(X'_{i_1\ldots i_\ell}M)) = v\left(\sum_{j=i_1}^{i_\ell} m_j\right)$$

y se cumple:

$$W(M) = \sum_{i=1}^{r} v(m_i) = \sum_{\substack{i=1 \\ i\notin\{i_1\ldots i_\ell\}}}^{r} v(m_i) + \sum_{j=i_1}^{i_\ell} v(m_j) =$$

$$= \sum_{\substack{i=1 \\ i\notin\{i_1,\ldots,i_\ell\}}}^{r} v(m_i) + v\left(\sum_{j=i_1}^{i_\ell} m_j\right) - v\left(\sum_{j=i_1}^{i_\ell} m_j\right) + \sum_{j=i_1}^{i_\ell} v(m_j)$$

$$= W(X_{i_1\ldots i_\ell}M) - W(M(X'_{i_1\ldots i_\ell}M)) + W(X'_{i_1\ldots i_\ell}M)$$

Luego $\phi(v) \in \mathcal{W}^*$.

<u>Lema 1.4.1.4</u>.- El conjunto $\gamma(\mathcal{V}) \subset \mathcal{W}^*$ queda caracterizado por el conjunto $\mathcal{C}$ donde $\mathcal{C} = \{C : \Delta \longrightarrow R^+ \cup \{0\}$, convexas$\}$.

<u>Demostración</u>.- Demostraremos que existe una correspondencia biunívoca $\Gamma$, entre $\mathcal{C}$ y $\mathcal{V}$

$$\Gamma : \mathcal{C} \longrightarrow \mathcal{V}$$

1) $\forall C \in \mathcal{C} \ \exists \ v \in \mathcal{V} \ / \ \Gamma(C) = v.$

$\forall x \in (R^+)^k$, sea $x = (x_1,\dots,x_k)$

Definimos:

$$\Gamma(C)(x) = C(\frac{x}{\Sigma x_i}) \cdot \Sigma x_i$$

$\Gamma(C) \in \mathcal{V}$, en efecto:

$\forall x,y \in (R^+)^k$

$$\Gamma(C)(x+y) = C(\frac{x+y}{\Sigma x+\Sigma y})(\Sigma x+\Sigma y) =$$

$$= (\Sigma x+\Sigma y)\ C(\frac{\Sigma x}{\Sigma x+\Sigma y} \cdot \frac{x}{\Sigma x} + \frac{\Sigma y}{\Sigma x+\Sigma y} \cdot \frac{y}{\Sigma y}) \leq$$

$$\leq \Sigma x\ C(\frac{x}{\Sigma x}) + \Sigma y\ C(\frac{y}{\Sigma y}) = \Gamma(C)(x) + \Gamma(C)(y)$$

2) $\forall v \in \mathcal{V}, \ \exists \ C \in \mathcal{C} \ / \ \Gamma^{-1}(v) = C$

$\forall x \in \Delta$ definimos:

$$\Gamma^{-1}(v)(x) = v(x)$$

$\Gamma^{-1}(v) \in \mathcal{C}$, en efecto:

$\forall \lambda \in [0,1]$, $\forall x,y \in \Delta$ se tiene:

$$\Gamma^{-1}(v)(\lambda x+(1-\lambda)y) = v(\lambda x + (1-\lambda)y) \leq$$

$$\leq v(\lambda x) + v((1-\lambda)y) = \lambda v(x) + (1-\lambda)v(y)$$

Con este resultado hemos conseguido caracterizar un subconjunto de $\mathcal{W}^*$ a través de las funciones convexas en $\Delta$ y nos permite enunciar el siguiente teorema

Teorema 1.4.1.1.- Para toda situación $(P,S)$ t.q. $S$ venga definida por [1.2.3], una condición suficiente para la existencia de una función bondad de tipo B1 es la existencia de una función convexa en $\Delta$, $C$ t.q.

$$B(S) = B((D_1,\ldots,D_r)) = \sum_{i=1}^{r} \left(\sum_{j=1}^{k} a_S(D_i,j)\right).C(p_{D_i}) \qquad [1.4.8]$$

Demostración.- Inmediata a partir de los lemas 1.4.1.3 y 1.4.1.4.

## 1.5. Dos casos particulares de función B1.

### 1.5.1. Función Bondad basada en la Entropía.

Si consideramos la función entropía, $H$, definida en $\Delta$

$$H(p) = -\sum_{j=1}^{k} p_j \log p_j$$

es cóncava en $\Delta$.

Tomamos como función bondad aquella definida en [1.4.8] siendo la función $C$:

$$C(p) = \log k - H(p), \quad \forall p \in \Delta$$

que es función convexa en $\Delta$, luego

$$\forall S \in A_g(P,S) \;/\; S = (D_1,\ldots,D_r)$$

$$B(S) = \sum_{i=1}^{r} \left(\sum_{j=1}^{k} a_S(D_i,j)\right).(\log k - H(P_{D_i}))$$

es una función bondad B1.

Se tiene:

$$B(s) = \sum_{i=1}^{r} \left(\sum_{j=1}^{k} a_S(D_i,_j)\right).(\log k + \sum_{j=1}^{r} (p_{D_i})_j \log(p_{D_i})_j)$$

El máximo valor, $M$ que puede tomar esta función es:

$$M = \log k \sum_{i=1}^{r} \sum_{j=1}^{k} a_S(D_i,j)$$

y por [1.2.3]

$$M = \log K \cdot \sum_{j=1}^{k} a_S(\Pi K\xi_i, j) = \log k \cdot \sum_{j=1}^{k} S(I_1, j) \qquad [1.5.1]$$

Este valor es alcanzado para aquellas situaciones $(P,S)$ tales que existe un $S \in A_g(P,S)$, $S = (D_1,\ldots,D_r)/r \leq k$ y $P_{D_i}$ es un vértice del simplex $\Delta$.

Este resultado no es soprendente pues era de esperar que la solución óptima tienda a dirigir los valores de sus segmentos asociados en los puntos de mínima ambigüedad, esto es, los vértices del simplex.

El mínimo valor, $m$, que puede tomar esta función es $m = 0$.

Este valor se alcanza en aquellas situaciones $(P,S)$ tales que existe un $S \in A_g(P,S)$, $S = (D_1,\ldots,D_r)/r = 1$ y $P_{D_i} = (\frac{1}{k}, \ldots, \frac{1}{k})$.

Estas situaciones corresponden a aquellas en que la función de semejanza cumple:

$$S(I_1,1) = S(I_1,2) = \ldots = S(I_1,k) \qquad [1.5.2]$$

Y la solución que da este valor para dicha situación es la partición constituida por un solo elemento, $\Pi K\xi_i$.

Este resultado, tiene también una interpretación muy intuitiva, pues establece que la situacion peor o con menos información inicial en ausencia de la información proporcionada por las variables explicativas, es la definida por [1.5.2], o que da la misma importancia a cada uno de los grupos; y que equivale a que el valor de esta solución en $\Delta$ sea el centro del simplex.

1.5.2. Función bondad basada en el máximo.

Si consideramos la función máximo en $\Delta$ definida por:

$$C(p) = \max_{1 \leq j \leq k} p_j, \quad \forall p \in \Delta \quad y \quad p = (p_1, \ldots, p_k)$$

Es trivial que $C \in V$, luego la función bondad definida por:

$$\forall S \in A_g(P,S) \;/\; S = (D_1, \ldots, D_r)$$

$$B(S) = \sum_{i=1}^{r} \left( \sum_{j=1}^{k} a_S(D_i, j) \right) . C(P_{D_i}) \text{ es de tipo B1.}$$

El máximo valor que toma $C$, es, análogamente a la función basada en la entropía, en los vértices del simplex $\Delta$. Siendo el máximo valor $M$ que toma $B$:

$$M = \sum_{i=1}^{r} \sum_{j=1}^{k} a_S(D_i; j) = \sum_{j=1}^{k} S(I_1, j) \qquad [1.5.3]$$

Que corresponde al mismo tipo de interpretación dada para la función definida en 1.5.1.

El mínimo, $m$, que toma esta función, viene dado por:

$$m = \frac{1}{k} \sum_{j=1}^{k} S(I_1, j) \qquad [1.5.4]$$

Que corresponde, análogamente a la función basada en la entropía, a aquellas situaciones donde se cumple [1.5.2].

## 1.6. Conclusiones

En este apartado vamos a exponer alguna de las ventajas que la solución dada al problema de discriminación, tal y como la hemos planteado, basada en la optimalidad de las funciones bondad B1, presenta, respecto de las soluciones dadas por la estadística clásica cuando suponemos que los datos corresponden a observaciones de las variables a través de una muestra aleatoria simple.

Consideremos la situación $(P,S)$ donde $S$ venga definido por medio de [1.2.3] a través de una función $\psi$ definida en [1.2.2].

Se tendrá:

$$\forall I' \quad I_1 \quad S(I',\ell) = \sum_{i \in J'} \psi(i,\ell) = |\{i \in I'/\eta(i) = \ell\}|$$

La solución, $S$, $(D_1,D_2,\ldots,D_r)$ viene definida por [1.2.10]

$$(P_{D_i})_\ell = \frac{|\{j \in J_1/\xi_1\ldots\xi_r(j) \in D_i,\ \ \eta(i) = \ell\}|}{|\{j \in J_1/\xi_1\ldots\xi_p(j) \in D_i\}|}$$

luego $P_{D_i}$ representa la distribución de la variable $\eta$ condicionada al segmento $D_i$, estimado a partir del conjunto de elementos observados $I_1$.

La solución bayesiana, en nuestro esquema vendrá dada por:

$$S(i_1,\ldots,i_p) \begin{cases} P_{\{i_1\ldots i_p\}} & \text{si} \quad \{j \in I_1/\xi_1,\ldots,\xi_p(j) = i_1,\ldots,i_p\} \neq \emptyset \\ (\frac{1}{k},\ldots,\frac{1}{k}) & \text{si} \quad \{j \in I_1/\xi_1,\ldots,\xi_p(j) = i_1,\ldots,i_p\} = \emptyset \end{cases}$$

y su clasificador asociado, $C_s$

$$C_s(i)=j_o \text{ si } \operatorname*{Max}_j (S(\xi_1,\ldots,\xi_p(i))_j = (S(\xi_1\ldots\xi_p(i))_{j_o}; \quad i \in I-I_1$$

La desventaja que tiene esta solución es la existencia de valores de $\Pi K \xi_i$ en donde no se han observado individuos. En este caso dicha solución no entraría dentro del conjunto de soluciones coherentes al problema $(P,S)$ y por tanto, evitaríamos en nuestro esquema el inconveniente que tiene la solución Bayesiana. El óptimo vendría dado en este caso por aquella partición entre las admisibles que maximiza

$$B(S) = \sum_{i=1}^{r} |\hat{I}_1(D_i)| C(P_{D_i}) = |I_1| \; E(C(P_{D_i})) \qquad [1.6.1]$$

es decir las funciones bondad en este caso vienen dadas como la esperanza matemática de $C(P_{D_i})$ respecto de la distribución marginal de la partición $(D_1,\ldots,D_r)$ multiplicada por el cardinal del conjunto de individuos etiquetados.

Cuando en todos los estados se han observado individuos, si hacemos

$$g = \min_{(i_1\ldots i_p)\in \Pi K\xi_i} |\hat{I}_1(\{i_1,\ldots,i_p\})|$$

Entonces la "solución bayes" coincide con la óptima en B1 y coinciden con $\{\{i_1,\ldots,i_p\},\ (i_1,\ldots,i_p) \in \Pi K\xi_i\} \in A_g(P,S)$.

Otra ventaja que presenta la solución propuesta frente a la clásica es que $g$ puede fijarse a priori; y por tanto podemos buscar la óptima entre las significativas.

# CAPITULO II

2.1. Capacidad descriptiva de las soluciones

La definición 1.1.2.1 de solución al problema de discriminación P quedaba justificada frente a otras posibles caracterizaciones de la solución debido a su poder descriptivo. Nos interesa estudiar como podemos explicitar esta capacidad descriptiva del tipo de soluciones que manejamos. Para ello introducimos los conceptos de hipótesis y de segmento compatible con una hipótesis.

A partir de estos elementos podremos contestar a preguntas a la solución acerca de su comportamiento respecto de determinadas hipótesis sobre los valores de $\eta$.

Definición 2.1.1.- Diremos que $H$ es una hipótesis de $P$ cuando $H \subset \Delta$.

Definición 2.1.2.- Sea $S \in A_g(P,S)$ y $D$ un segmento de $S$, diremos que $D$ es compatible con la hipótesis $H$ de $P$ cuando $p_D \in H$.

Estas dos definiciones nos permitirán establecer aquellos subconjuntos de $\Pi K\xi_i$ que favorecen determinadas hipótesis sobre los valores de la variable grupo. Así podemos estar interesados en comprobar si una solución define algún estado de $\eta$, dicha pregunta a la solución $S$ la podemos formular en los siguientes términos:

Dada $S \in A_g(P,S)$, ¿existe algún segmento $D$ de $S$ que sea compatible con la hipótesis $H_i = \{(0,0,\ldots,1^{(i},0,\ldots,0)\} \subset \Delta$?.

Definición 2.1.3.- Sea $S \in A_g(P,S)$, diremos que $D \subset \Pi K\xi_i$, define el estado $j$ de $\eta$ cuando $D$ es un subconjunto de un segmento com patible con $H_j = \{(0,\dots,1^{(j},0,\dots,0)\} \subset \Delta$.

Diremos que $S \in A_g(P,S)$, define el estado $j$ de $\eta$ cuando algún segmento de $S$ define el estado $j$ de $\eta$.

Teorema 2.1.1. Sean $S_1, S_2 \in A_g(P,S)$, si $S_1 \lessapprox S_2$ y $S_1$ define el estado $j$ de $\eta$ entonces $S_2$ define el mismo estado de $\eta$.

Demostración.- Sea $D$, segmento de $S_1$ que define el estado $j$ de $\eta \iff S_1^{-1}(\{0,\dots,1^{(j},0,\dots,0\}) = D \iff S(D,\ell) = 0 \quad \forall \ell \neq j \iff$ $\iff \forall D' \subset D,\ S(D',\ell) = 0 \implies \forall S \in A_g(P,S)$ si $D' \subset D$ es segmento de $S$ y $S_1 \lessapprox S$ entonces $D' = D \implies D$ es segmento de $S_2$ y $P_D = \{0,\dots,1^{(i},\dots,0\} \implies S_2$ define el estado $j$ de $\eta$.

Este resultado válido para funciones bondad de tipo B1, será deseable que lo cumplan todas las funciones bondad que generemos, y por tanto lo utilizaremos más adelante como una propiedad a comprobar a partir de los diferentes criterios de bondad que introduzcamos.

Definición 2.1.4.- Diremos que $S \in A_g(P,S)/S = (D_1, D_2, \dots, D_r)$, define totalmente el estado $j$ de $\eta$ si $\exists\ i \in \{1,2,\dots,r\}/$ $/S^{-1}(\{0,\dots,1^{(j},0,\dots,0\}) = D_i$ y

$$P_{D_n} \in \{(p_1,p_2,\dots,p_k) \in \Delta \quad \text{y} \quad P_j = 0\} \qquad \forall n \neq i.$$

Consecuencias inmediatas de la definición 2.1.4. son las siguientes propiedades:

P.1 Si $S \in A_g(P,S)$ define totalmente el estado $j$ de $\eta$ entonces $S$ define el estado $j$ de $\eta$.

P.2 Si $S_1 \in A_g(P,S)$ define totalmente el estado $j$ de $\eta$ entonces $\forall S_2 \in A_g(P,S)$ / $S_2 \succsim S_1$ se cumple que $S_2$ define el mismo estado $j$ de $\eta$.

<u>Demostración de P.2</u>.- Por P.1 y el teorema 2.1.1 se tiene que $S_2$ define el estado $j$ de $\eta$.

Sea $S_2 = (D_1, D_2, \ldots, D_r)$ y supongamos que $D_j$ es el segmento que define el estado $j$ de $\eta$ y además $D_{j_1}$ es segmento de $S_1$; se tiene:

$\forall i \in \{1,2,\ldots,r\}$; $i \neq j_1$, $\exists$ un segmento $D'$ de $S_1$ y $D' \neq D_{j_1} / D_i \subset D'$.

Si $D'$ es segmento de $S_2$ se cumple que $a_S(D',j_1) = 0$ luego $a_S(D_i,j_1) = 0$, por tanto $P_{D_i} \in \{(p_1,\ldots,p_k) \in \Delta$ y $p_j = 0\}$ luego $S_2$ define totalmente el estado de $j$ de $\eta$.

<u>Definición 2.1.5</u>.- Diremos que $S \in A_g(P,S)$ define totalmente $\eta$ cuando todos los segmentos de $S$ definen algún estado de $\eta$.

A partir de la definición 2.1.5. tenemos las siguientes propiedades:

P.1 Si $S \in A_g(P,S)$ define totalmente $\eta$ entonces $\tau(S) \leq k$

P.2 Si $S \in A_g(P,S)$ define totalmente a $\eta$ entonces $\forall S' \succsim S$, $S'$ define totalmente a $\eta$.

Luego si existe una solución admisible que defina totalmente a $\eta$ se tendrá que es la solución óptima para el conjunto de soluciones admisibles relacionadas con dicha solución a través de la relacción de orden parcial $\lesssim$ .

Estos resultados nos muestran la importancia que tiene el considerar como hipótesis a estudiar los vértices del ximplex $\Delta$. Dado que en general las situaciones a las que nos enfrentamos no admiten una solución que cumpla con la definición 2.1.5, dirigiremos nuestra atención a hipótesis menos restrictivas que los vértices.

Introducimos una función de desemejanza d, en el simplex que nos va a servir para establecer un método que nos permita relajar la hipótesis que fijemos sobre $\Delta$

$$d : \Delta \times \Delta \longrightarrow R^{+} \cup \{0\}$$

$$\begin{array}{ll} P.1 & d(x,y) = d(y,x) \\ P.2 & d(x,y) = 0 \iff x = y \end{array} \qquad [2.1.1]$$

En particular podremos tomar como d, una métrica definida en $\Delta$.

Procedimiento para relajar hipótesis

Consideremos $H \subset \Delta$, hipótesis para $(P,S)$.

Supondremos que H viene definida por un punto de $\Delta$.

Sea $S \in A_g(P,S)$ representamos por $S^{-1}(H)$ los segmentos de S compatibles con H. Sí $S^{-1}(H) = \emptyset$, construimos una hipótesis $H' \supset H$ t.q. $S^{-1}(H') \neq \emptyset$.

Sea $H^{\ell} = \{p \in \Delta \,/ d(p,H) \leq \ell\}$, $\ell \in R^{+}$ [2.1.2]

Calculamos:

$$\ell_o = \min \ \{\ell : S^{-1}(H\ ) \neq \emptyset\} \qquad [2.1.3]$$

Teniendo que $H^{\ell_o}$ es la hipótesis más cercana a H de forma que exista al menos un segmento de S que sea compatible con ella. Denominaremos por $H^{\ell_o}$ la hipótesis relajada de H a través de la solución S.

Fuerza de una hipótesis

Definición 2.1.6.- Sea $S \in A_g(P,S)$ y H una hipótesis para (P,S), denominaremos fuerza de una hipótesis H a través de S y lo designaremos por $F_S(H)$ a la suma de los pesos de los segmentos compatibles con ella a través de la solución S.

A partir de la definición 2.1.6 son inmediatas las siguientes propiedades:

P.1 $\forall S \in A_g(P,S)$, la fuerza $F_S$ es una función:

$$F_S : P(\Delta) \longrightarrow R^+ \cup \{0\}$$

$$H \longrightarrow F_S(H) = f(S^{-1}(H))$$

P.2 $\forall S \in A_g(P,S)$

$\forall H_1,H_2 \subset \Delta \ / \ H_1 \subset H_2$

Se cumple que $F_S(H_1) \leq F_S(H_2)$

P.3 La condición necesaria y suficiente para que una solución $S \in A_g(P,S)$ tenga algún segmento compatible con una hipótesis H es que $F_S(H) \neq 0$.

Ambigüedad de una Hipótesis.

Definición 2.1.7.- Sea $S \in A_g(P,S)$, y sea $H$ una hipótesis de $(P,S)$ que supondremos que es un punto de $\Delta$ y sea $D_j$ un segmento de $S$, definimos ambigüedad de $H$ por $D_j$ y la designaremos por $A_{D_j}(H)$, como la desemejanza de $P_{D_j}$ a $H$

$$A_{D_j}(H) = d(P_{D_j},H) \qquad [2.1.4]$$

Aunque la anterior definición de ambigüedad de una hipótesis por un segmento solo es válida para hipótesis de $(P,S)$ que puedan expresarse como puntos de $\Delta$, siempre es posible generalizarla a cualquier tipo de hipótesis, bastaría con definir una desemejanza de punto a conjunto en $\Delta$ a partir de la desemejanza definida en [2.1.1].

A partir de las definiciones 2.1.7 y 2.1.8 se tienen las siguientes propiedades:

P.1 Sea $S \in A_g(P,S)$ y $D$ un segmento de $S$.
Sea $H$ una hipótesis para $(P,S)$.
$D$ es compatible con $H \iff A_D(H) = 0$

P.2 Sea $S \in A_g(P,S)$,, $S = (D_1,D_2,\ldots,D_r)$.
Sea $H$ una hipótesis de $(P,S)$.
Si $H^{\ell_o}$ es la hipótesis relajada de $H$ a través de $S$, definida en [2.1.2] y [2.1.3] entonces $\ell_o = \min\limits_{i\in\{1,\ldots,r\}} A_{D_i}(H)$

P.3 Sea $S \in A_g(P,S)$,
Sea $H$ una hipótesis de $(P,S)$.

$$F_S(H) \neq 0 \iff \exists \text{ un segmento } D \text{ de } S \;/\; A_D(H) = 0$$

P.4 $\forall H_1, H_2 \subset \Delta$, $\forall S \in A_g(P,S)$

Si $H_1 \subset H_2 \implies A_{D_i}(H_1) \geq A_{D_i}(H_2)$ $\forall D_i$ segmento de S

Definición 2.1.8.- Sea $S \in A_g(P,S)$ y H hipótesis de (P,S), definimos ambigüedad de H por S y lo designamos por $A_S(H)$ a:

$$A_S(H) = h(A_{D_1}(H),\ldots,A_{D_\ell}(H)), \quad S = (D_1,\ldots,D_\ell)$$

donde cada h es una función monótona creciente en cada componente y siendo

$$h(0) = 0$$

Un caso particular importante y que consideraremos más adelante es:

$$h(A_{D_1}(H),\ldots,A_{D_\ell}(H)) = \Sigma\, a_i\, A_{D_i}(H), \quad a_i \in R^+ \cup \{0\} \qquad [2.1.5]$$

En la definición anterior, hemos utilizado la notación h para designar a la familia de funciones definidas de $R^\ell$ en R $\forall \ell \in N$ y $\ell < \infty$, que caracterizan la ambigüedad de soluciones de tamaño $\ell$.

En lo que sigue nos referiremos a esta familia de funciones como función h.

Definición 2.1.9.- Sea $S \in A_g(P,S)$, y sea H una hipótesis de (P,S). Diremos que S define la hipóesis H si y sólo si $A_S(H) = 0$.

Supongamos que $S$ viene definida a través de $\psi$ por [1.2.3].

$\forall \ell \in K\eta$, obtenemos $(P_\ell,S_\ell)$ a partir de (P,S) definiendo sus componentes:

<u>Definición 2.1.10</u>.- Sea $P = (I,\eta,E)$ definimos $P = (I_\ell,\eta_\ell,E_\ell)$ donde:

$I_\ell = I$

$E_\ell = E$

$\eta_\ell$ es una variable categórica que se deduce de $\eta$ solapando en una única categoria todas las categorías de $\eta$ salvo la "$\ell$" que a su vez define la 2 categoría de $\eta_\ell$.

<u>Definición 2.1.11</u>.- Sea $S$ asociada a P, y sea $P_\ell$ cumpliendo con definición 2.1.10 definimos $S_\ell$ asociada a $P_\ell$:

$$S_\ell(I',1) = \sum_{\substack{j=1 \\ j\neq\ell}}^{r} S(I',j) \qquad \forall I' \subset I$$

$$S_\ell(I',2) = S(I',\ell) \qquad \forall I' \subset I$$

Si en $K\eta$, consideramos la partición $K'\eta$, se pueden obtener los subproblemas $(P_{K'\eta},S_{S'\eta})$ a partir de $(P,S)$ definiendo sus componentes:

<u>Definición 2.1.12</u>.- Sea $P = (I,\eta,E)$, definimos $P_{K'\eta} = (I_{K'\eta},S_{K'\eta}, E'_{K'\eta})$ donde:

$$I_{K'\eta} = I$$

$$E_{K'\eta} = E$$

$\eta_{K'_\eta}$ es una variable con $|K'_\eta|$ categorias que se deduce de en la siguiente forma:

Toda categoría de $\eta_{K'_\eta}$ se deduce de solapar las categorias de que pertenecen al mismo conjunto de la partición $K'_\eta$.

Definición 2.1.13.- Sea $S$, asociada a $P$ y sea $P_{K'_\eta}$ cumpliendo con la definición 2.1.12, definimos $S_{K'_\eta}$ asociada a $P_{k'_\eta}$ :

$$S_{K'_\eta}(I',\ell) = \sum_{j\in L} S(I',j), \quad \forall I' \subset I, \quad \forall \ell \in \eta_{K'_\eta}$$

donde $L \subset K\eta$ es el elemento de $K'\eta$ que define la $\ell$-ésima categoría de $\eta_{K'_\eta}$.

Son inmediatos los siguientes resultados:

R1 $\quad P_\ell \equiv P_{\{K\eta-\{\ell\},\{\ell\}\}}$

R2 $\quad P \equiv P_{\{\{1\},\{2\},\dots,\{k\}\}}$.

R3 Teorema 2.1.2.- $\exists$ $S \in A_g(P,S)$, que define totalmente el estado $j$ de $\eta$ si y sólo si $\exists$ $S_j \in A_g(P_j,S_j)$ tal que $A_{S_j}(H) = 0$ para $H = \{(0,1),\ (1,0)\}$.

Demostración.- Es condición necesaria:

Sea $S \in A_g(P,S)$, $S = (D_1,D_2,\dots,D_r)$ cumpliendo con definición 2.1.4.

Sabemos que $\forall y \in \prod_{i=1}^{p} K\xi_i$, $\exists \ell \in \{1,2,\dots,r\}$ / $y \in D_\ell$.

Definimos $S_j : \prod K\xi_i \longrightarrow \Delta$ y $\Delta \subset R^2$, donde $S_j(y)$ queda definido a través de [1.2.9] estando $S_j$ definida por [2.1.6].

Sea $H = \{(0,1),\ (1,0)\}$, se cumple que $A_S(H) = 0$.

Si $y \in D_\ell$, como $S$ definte totalmente el estado $j$, de la definición 2.1.4 $P_{D_\ell} = (0,\dots,1^{(j},0,\dots,0)$ ó bien $P_{D_\ell} = (r_1,\dots,0,\dots,r_k) \in \Delta,$, $\Delta \subset R^k$.

Por tanto $S_j(y) = (0,1)$ ó bien $S_j(y) = (1,0)$.

Luego $\forall y \in \Pi K\xi_i \quad S_j(y) \in H$ y además $S_j$ tiene tamaño 2.

Sean $(D_1^j, D_2^j) = S_j$ t.q. $P_{D_1^j} = (0,1)$, $P_{D_2^j} = (1,0)$.

Se tiene: $A_{D_i^j}(H) = 0$, $i=1,2$, luego $A_{S_j}(H) = 0$.

Es condición suficiente:

Sea $S_j \in A_g(P_j, S_j)$ t.q. $A_{S_j}(H) = 0$.

$A_{S_j}(H) = 0 \Longrightarrow S_j$ es de tamaño 2, sea $S_j = (D_1, D_2)$ se cumple:

$P_{D_1} = (0,1)$; $P_{D_2} = (1,0) \Longrightarrow$ la solución de $(P,S)$, asociada a $(D_1, D_2)$, cumple con la definición 2.1.4.

En efecto:

Si $S = (D_1, D_2)$ es admisible entonces $S$ viene definida a partir de [1.2.10] y se tiene:

$$(P_{D_1})_j = \frac{a_S(D_1, j)}{\sum_{\ell=1}^{k} a_S(D_1, \ell)} = \frac{a_{S_j}(D_1, 2)}{\sum_{\ell=1}^{2} a_{S_j}(D_1, \ell)} = 1$$

$$(P_{D_2})_j = \frac{a_S(D_2, j)}{\sum_{\ell=1}^{k} a_S(D_2, \ell)} = \frac{a_{S_j}(D_2, 2)}{\sum_{\ell=1}^{2} a_{S_j}(D_2, \ell)} = 0$$

que nos dice que $S = (D_1, D_2) \in A_g(P,S)$ y cumple con la definición 2.1.4.

De forma análoga demostraríamos el siguiente resultado:

<u>Teorema 2.1.3</u>.- Existe $S \in A_g(P,S)$ que define totalmente los estados $i_1, i_2, \ldots, i_r$ de $\eta$ si y sólo si

$\exists\; S' \in A_g(P_{\{\{i_1\},\dots,\{i_r\},K\eta-\{i_1,\dots,i_r\}\}},\; S_{\{\{i_1\},\dots,\{i_r\},K\eta-}$
$_{-\{i_1,\dots,i_r\}\}}$ tal que $A_S(H) = 0$ para
$H = \{(0,\dots,1),\ (0,\dots,1,0),\overset{(r}{\dots},(1,0,\dots,0)\}$.

Corolario 2.1.1.- Dada $(P,S)$; $S \in A_g(P,S)$ define totalmente a $\eta \iff A_s(H) = 0$ para $H = \{(0,1,\dots,1),(0,\dots,1,0),\overset{k}{\dots},(1,0,\dots,0)\}$.

La demostración es inmediata a partir del teorema 2.1.3.

2.2. Funciones Bondad basadas en la ambigüedad.

Definición 2.2.1.- Dada $(P,S)$, siendo $d$ una desemejanza en $\Delta$ cumpliendo con [2.1.1] y $B$ una función bondad para $A_g(P,S)$, diremos que $B$ es compatible con la hipótesis $H$ de $(P,S)$ si se cumple:

$$\forall S_1,S_2 \in A_g(P,S)/A_{S_1}(H) < A_{S_2}(H) \iff B(S_1) > B(S_2) \qquad [2.2.1]$$

La definición 2.2.1. nos vá a permitir caracterizar las funciones bondad y por tanto la solución óptima al problema, condicionado a la hipótesis que fijemos a priori sobre los estados de la variable dependiente $\eta$.

En general la hipótesis que se fijará vendrá definida por los vértices del simplex $\Delta$, debido a que, caso de que exista alguna solución al problema que defina totalmente los estados de $\eta$, esta será óptima para toda función bondad que cumpla con la condición 2.2.1. Esta hipótesis la designaremos por $H_{1,2,\dots,k}$.

No siempre, será $H_{1,2,\dots,k}$ la hipótesis que fijaremos, dado que esta dependerá de las conclusiones que sobre la situación estudiada pretendamos sacar. La ventaja que presenta trabajar con funciones bondad que cumplan con la definición 2.2.1 reside en la versatilidad que tiene respecto de los objetivos que estudiemos y que vienen expresados mediante las hipótesis que pretendemos contrastar. Así, nos podrá interesar obtener aquellos segmentos de la población que favorecen un determinado valor de la variable dependiente $\eta$.

Supongamos que sea la categoría i-ésima de $\eta$ sobre la que estamos interesados. Para obtener la solución a dicho problema consideraremos el problema $(P_i,S_i)$ definido a partir de la definición 2.1.10 y [2.1.6] y sobre este problema la hipótesis $H_{12}$. La solución que buscamos vendrá dada como aquella que optimice una función bondad compatible con la hipótesis $H_{12}$.

Una aproximación descriptiva al problema $(P,S)$ consiste en obtener las soluciones a los problemas $(P_i,S_i)$ $i=1,\dots,k$ como indicamos en el párrafo anterior. Del análisis de estas soluciones obtendríamos qué segmentos son los que favorecen cada una de las categorias de la variable dependiente aunque el conjunto de estos no nos darían una solución que pertenezca a $A_g(P,S)$.

A veces interesará estudiar cuáles son los segmentos que favorecen una determinada categoría de $\eta$ en detrimento de otra categoría, estos serán los segmentos que favorecen la oposición entre dos categorias. Para determinar la solución que aplicar dicha oposición, definimos $(P_{1\leftrightarrow 2},S_{1\leftrightarrow 2})$ a partir de $(P,S)$, suponiendo que queremos estudiar la oposición entre las categorias 1 y 2.

Sean

$$P_{1\leftrightarrow 2} = (I_{1\leftrightarrow 2}, \eta_{1\leftrightarrow 2}, E_{1\leftrightarrow 2})$$

$$P = (I, \eta, E)$$

Donde:

$$I_{1\leftrightarrow 2} = \{i \in I \ / \ \eta(i) \in \{1,2\}\}$$

$\eta_{1\leftrightarrow 2}$, queda definida por las categorías 1,2 de $\eta$

$$E_{1\leftrightarrow 2} = E$$

Para definir $S_{1\leftrightarrow 2}$, supondremos que $S$ viene definida a partir de [1.2.3] mediante una función $\psi$ definida en [1.2.2].

Definimos:

$$S_{1\leftrightarrow 2}(I',\ell) = \sum_{i\in I'} \psi(i,\ell); \qquad \begin{array}{l} \ell = 1,2 \\ I' \subset I_{1\leftrightarrow 2} \end{array}$$

Considerando $H_{1,2}$ para $(P_{1\leftrightarrow 2}, S_{1\leftrightarrow 2})$, la solución que explica la oposición entre las clases 1 y 2 vendrá dada como aquella que maximice una función bondad que cumpla con la definición 2.2.1 para $(P_{1\leftrightarrow 2}, S_{1\leftrightarrow 2})$ y $H_{1,2}$.

De forma análoga obtendríamos la solución que explique la oposición entre 2 conjuntos de categorías de la variable $\eta$. Y de forma más general la oposición de más de 2 conjuntos de categorias de $\eta$ entre sí.

La obtención de las soluciones que determinen oposiciones nos va a permitir definir una función de desemejanza entre las categorias de la variable $\eta$ en la siguiente forma:

Definición 2.2.2.- $\forall S \in A_g(P,S)$ y para toda función bondad definida en $A_g(P,S)$, definimos poder discriminante de $S$, y lo designamos $PD_B(S)$ a:

$$PD_B(S) = \frac{B(S) - m_B}{M_B - m_B} \qquad [2.2.2]$$

donde:

$M_B$ representa el máximo valor que puede tomar la función bondad $B$.

$m_B$ representa el mínimo valor que puede tomar la función $B$.

Definición 2.2.3.- $\forall (P,S)$ y para toda función bondad definida en $A_g(P,S)$, definimos poder discriminante de $(P,S)$ y lo designamos por $PD_B$ a:

$$PD_B = \frac{M^* - m_B}{M_B - m_B} \qquad [2.2.3]$$

donde:

$M^* = \max_{S \in A_g(P,S)} B(S)$ ó en su defecto el valor de la función bondad para la solución que tomemos como óptima.

De las definiciones anteriores se deduce inmediatamente:

P.1 $\forall S \in A_g(P,S)$ y para toda función bondad $B$ definida en $A_g(P,S)$. Se cumple:

$$0 \leq PD_B(S) \leq 1$$

P.2 $\forall (P,S)$ y para toda función bondad $B$ definida en $A_g(P,S)$ se cumple:

$$0 \leq PD_B \leq 1$$

Dado que el poder discriminante de una solución depende de la función bondad sobre la que está basada, la principal ventaja que presenta trabajar con poderes discriminantes en vez de con funciones bondad es que en estos sabemos su rango de variación. Esta propiedad resulta particularmente interesante cuando consideramos poderes discriminantes de problemas o situaciones distintas pues nos va a permitir su comparación.

<u>Definición 2.2.4</u>.- Sea $(P,S)$, consideramos las categorias $i$ y $j$ de la variable $\eta$ de $P$ definimos la desemejanza entre $i$ y $j$ y lo designamos por $\delta_B(i,j)$ a:

$$\delta_B(i,j) = PD_B(P_{i\leftrightarrow j},S_{i\leftrightarrow j})$$

donde $B$ es una función bondad para $(P_{i\leftrightarrow j},S_{i\leftrightarrow j})$ compatible con la hipótesis $H_{1,2}$.

Si $\delta_B(i,j) = 0$ quiere decir que las clases $i,j$ tienen una oposición mínima y vamos a poder identificarlas dependiendo de la función bondad que tomemos.

La desemejanza dada por la definición 2.2.4 nos va a permitir realizar un proceso de clustering sobre las categorias de la variable $\eta$ en la siguiente forma:

En la primera etapa las categorias 'i' y 'j' de $\eta$ se solapan en una única categoría que designaremos por $t$; si se cumple:

$$\delta_B(i,j) = \min_{\ell,s\in K\eta} \delta_B(\ell,s)$$

redefiniendo la desemejanza entre las nuevas categorías por:

$$\delta_B(K,\ell) = PD_B(P_{\{i,j\}\leftrightarrow\ell}, S_{\{i,j\}\leftrightarrow\ell}) \quad \forall \ell \in K\eta, \quad \ell \neq i, \quad \ell \neq j$$

y es fácil comprobar que coincide con la desemejanza definida en 2.2.4 para el problema $(P_{\{\{i,j\},\{1\},\{2\},\ldots,\{k\}\}}, S_{\{\{i,j\},\{1\},\{2\},\ldots,\{k\}\}})$, a este problema lo denominaremos problema generado en la etapa 1.

En etapas sucesivas se procede como en la primera etapa sobre el problema generado en la etapa anterior.

El proceso concluye en la etapa k-1 donde el problema generado en dicha etapa la variable dependiente contiene una única categoría.

El objetivo principal del proceso de Cluster sobre las categorías descrito más arriba es poder reducir el número de categorías de la variable dependiente con la consiguiente simplificación del problema de estimar esta a partir de E, además de hacer un estudio descriptivo sobre las similitudes entre las diferentes categorías de $\eta$ basándose en como E explica a estas.

Una forma sencilla de obtener las funciones bondad que cumplan con la definición 2.2.1 es encontrar una función real estrictamente decreciente, este tipo de caracterización nos viene dada en el siguiente resultado

<u>Teorema 2.2.1</u>.- Dada (P,S), si $\forall S \in \mathcal{A}_g(P,S)$ tenemos definida $A_S(H)$ para H, hipótesis de (P,S), que cumple con la definición 2.1.8 si $\exists$ una función $\nabla$ real de variable real monótona decreciente estrictamente entonces $\nabla(A_S(H))$ es una función bondad com

patible con H.

Demostración: Es inmediata a partir de la definición 2.2.1.

## 2.3. Funciones basadas en la ambiguedad y que sean Bl

Un conjunto particularmente interesante entre las funciones bondad compatibles con alguna hipótesis de $(P,S)$ son aquellas que además sean Bl.

Consideraremos la hipótesis $H_{12\ldots k} = H$ y sea $d$ la desemejanza que genera la ambiguedad de $H$; sea:

$$M = \max_{p\in\Delta} \min_{1<i<k} d(p,V_i); \quad V_i \text{ es vértice } \Delta \qquad [2.3.1]$$

Se cumple el siguiente resultado:

Teorema 2.3.1.- $\forall S \in A_g(P,S)$, $S = (D_1,D_2,\ldots,D\ )$, $A_S(H) = h(A_{D_1}(H),\ldots,A_D\ (H))$. Si es $h$ una combinación convexa de sus componentes entonces la función bondad $B(S) = h(M-A_{D_1}(H),\ldots$ $\ldots,M-A_{D_\ell}(H))$ es compatible con $H$.

Demostración: Sean $S_1,S_2 \in A_g(P,S)$ t.q. $S_1 = (D_1^1,\ldots,D_\ell^1)$, $S_2 = (D_1^2,\ldots,D_{\ell_2}^2)$, veremos:

$$A_{S_1}(H) < A_{S_2}(H) \iff B(S_1) > B(S_2)$$

$\Rightarrow$ )

$$B(S_1) = h(M-A_{D_1^1}(H),\ldots,M-A_{D_{\ell_1}^1}(H)) = \sum_{i=1}^{\ell_1} a_i^1(M-A_{D_\ell^1}(H)) =$$

$$= M \sum_{i=1}^{\ell_1} a_i^1 - \Sigma\, a_i^1 A_{D_i^1}(H) = M - h(A_{D_1^1}(H),\dots,A_{D_{\ell_1}^1}(H)) >$$

$$> M - h(A_{D_1^2}(H),\dots,A_{D_{\ell_2}^2}(H)) = M \sum_{i=1}^{\ell_2} a_i^2 - \sum_{i=1}^{\ell_2} a_2^i A_{D_i^2}(H) =$$

$$= h(M-A_{D_1^2}(H),\dots,M-A_{D_{\ell_2}^2}(H)) = B(S_2)$$

$\Leftarrow$) es análoga a $\Rightarrow$ ).

<u>Teorema 2.3.2</u>.- Para cualquier $(P,S)$ tal que $S$ venga definida a través de [1.2.3], sea la hipótesis $H_{12\dots k}$, y $d$ la desemejanza definida en $\Delta$ que genera la ambiguedad de la hipótesis $H_{12\dots k}$, si se cumple que $d(p,H)$ es concava respecto de $p$, entonces se puede construir una función $h$ que genera según el teorema 2.3.1 una bondad de tipo B1.

<u>Demostración</u>.- Sea $S \in A_g(P,S)$ y $S = (D_1,D_2,\dots,D\ )$

$$A_S(H) = h(A_{D_1}(H),\dots,A_{D_\ell}(H)) = \sum_{i=1}^{\ell} r_i\, A_{D_i}(H)$$

con

$$r_i = \frac{\sum_{j=1}^{k} a_S(D_i,j)}{\sum_{i=1}^{\ell}\sum_{j=1}^{k} a_S(D_i,j)} \qquad [2.3.2]$$

entonces:

$$B(S) = \sum_{i=1}^{\ell} r_i(M - A_{D_i}(H)) \qquad [2.3.3]$$

donde $M$ viene definido por [2.3.1].

Evidentemente cumple con las hipótesis del teorema 2.3.1 y como consecuencia la función bondad definida es compatible con H.

Veamos que es de tipo Bl:

Por el teorema 1.4.1.1, sabemos que cuando $S$ viene definida por [1.2.3] las funciones representadas por [1.4.8] son funciones Bl, además si $\exists$ B que sea Bl, $\forall L \in R^+$ se tiene que LB sigue siendo Bl, es decir:

$$L \cdot \sum_{i=1}^{\ell} \left( \sum_{j=1}^{k} a_S(D_i,j) \right) \cdot C(p_{D_i}) \qquad [2.3.4]$$

es Bl, para C convexa en $\Delta$ y $L \in R^+$.

Si tomamos $L = \left( \sum_{i=1}^{\ell} \sum_{j=1}^{k} a_S(D_i,j) \right)^{-1}$

que es una constante real positiva para toda situación (P,$S$) y como $C(P) = M - d(p,H)$ que es convexa en $\Delta$ por ser $d(p,H)$ concava en $\Delta$, se tiene que [2.3.4] coincide con [2.3.3] y por tanto es compatible con H.

Una particularización importante del resultado del teorema 2.3.2, se tiene cuando tomamos como función $\psi$, generadora de la función $S$, la definida por [1.2.1], pues las $a_i$ definidas por [2.3.2] coinciden con los pesos de los segmentos $D_i$ cuando la función peso viene definida por [1.2.12], luego la bondad puede expresarse como:

$$B(S) = \sum_{i=1}^{\ell} f(D_i) \cdot (M - A_{D_i}(H)) \qquad [2.3.5]$$

La importancia del resultado del teorema 2.3.2 y en particular de la expresión [2.3.5] radica en las restricciones que se imponen a la caracterización de la ambigüedad a la hora de generar funciones bondad que sean compatibles con $H_{12...k}$ y al mismo tiempo sean Bl. Recíprocamente a este resultado vamos a establecer cuando podemos generar una función de ambigüedad a través de una función bondad Bl. de forma que dicha función bondad sea compatible con la hipótesis $H_{12...k}$ a través de la ambigüedad generada.

Corolario al Teorema 2.3.2.- Sea $C$ convexa en $\Delta$ con valor constante $K_1$ en los vértices $v$ de $\Delta$ y $d(p,H) = K_1 - C(p)$, entonces la función bondad definida como en [2.3.4] es compatible con $H_{12...k}$.

Demostración.- Es trivial ver que $d(p,H)$ está bien definida para $H = H_{12...k}$ y $d(p,H)$ es cóncava en $\Delta$ pues $C$ es convexa. Luego estamos en las hipótesis del teorema 2.3.2 y por tanto la función bondad definida por [2.3.4] es compatible con $H_{12...k}$.

Un caso particular interesante se tiene cuando la función $C \in \mathcal{C}$ viene definida a partir de una distancia sobre $\Delta \subset R^k$.

Sea $\delta$, una métrica definida en $\Delta$ que cumple:

1) $\delta(c,v) = cte.$ $\forall v$, vértice de $\Delta$.

siendo $c$ el centro de 1; $c = (1/k,...,1/k)$ [2.3.6]

2) $\delta(c,p)$ es función convexa de $p$.

Si definimos la función:

$$C : \Delta \longrightarrow R$$

$$C(p) = \delta(c,p)$$

$C$ cumple con las hipótesis del corolario al teorema 2.3.2 y aplicando la caracterización dada por [2.3.5] tenemos:

$$B(S) = \sum_{i=1}^{\ell} f(D_i)\delta(p_{D_i},c) \quad \forall S \in A_g(P,S), \quad S = (D_1,\ldots,D_\ell) \qquad [2.3.7]$$

es una función bondad B1 compatible con $H_{12\ldots k}$ siempre que la función peso esté definida por [1.1.12] y estemos en las condiciones del teorema 2.3.2.

## 2.4. Casos particulares de funciones bondad B1 compatibles con $H_{12\ldots k}$

### 2.4.1. Ambigüedad basada en la Entropía.

Si tomamos como función $C$, la definida en 1.5.1, es trivial comprobar que $C$ cumple con las condiciones del corolario al teorema 2.3.2 luego $\forall D \subset \Pi K\xi_i \ / \ f(D) \geq g$, definimos ambigüedad de $D$ como:

$$A_D(H) = -\sum_{i=1}^{k} (p_D)_i \log(p_D)_i, \quad H = H_{12\ldots k} \qquad [2.4.1]$$

que hace que la función bondad definida en 1.5.1 sea compatible con $H_{12\ldots k}$.

### 2.4.2. Ambigüedad basada en la función máximo.

Tomando como función $C$, la definida en el apartado 1.5.2, es fácil comprobar que dicha función cumple con las hipótesis del corolario al teorema 2.3.2 luego $\forall D \subset \Pi K\xi_i \ / \ f(D) \geq g$ definimos la función ambigüedad como:

$$A_D(H) = 1 - \max_{1\leq j\leq k} ((p_D)_j), \quad H = H_{12\ldots k} \qquad [2.4.2]$$

que hace que la función bondad definida en 1.5.2 sea compatible con $H_{12\ldots k}$.

2.4.3. Ambigüedad Euclídea.

Si tomamos como función C la definida por:

$$C(p) = \sum_{i=1}^{k} (p_i - \frac{1}{k})^2, \quad p \in \Delta$$

esta cumple con las hipótesis al corolario del teorema 2.3.2.

En efecto:

1) $C \in \mathcal{C}$

2) $C(v) = \frac{k-1}{k} \quad \forall v$, vértice de $\Delta$.

Luego $\forall D \subset \Pi K\xi_i \;/\; f(D) \geq g$, definimos:

$$A_D(H) = \left(\frac{k-1}{k}\right)^2 - \sum_{i=1}^{k} (p_i - \frac{1}{k})^2 = \frac{k-1}{k} - (\Sigma\, p_i^2 - \frac{1}{k}) = 1 - \Sigma\, p_i^2$$

y por el teorema 2.3.2, definimos la función bondad B(S)

$$B(S) = \sum_{i=1}^{\ell} a_i \left(\sum_{j=1}^{k} (p_j - \frac{1}{k})^2\right) \quad S \in A_g(P,S)/S=(D_1,\ldots,D\ ) \qquad [2.4.3]$$

donde los $a_i$ vienen definidos por [2.3.2], que es Bl y compatible con $H_{12\ldots k}$.

Operando en [2.4.3] se tiene:

$$B(S) = \sum_{i=1}^{r} a_i \left(\sum_{j=1}^{k} p_j^2\right) - \frac{1}{k}$$

que respecto de las soluciones tiene el mismo comportamiento que la función:

$$B(S) = \sum_{i=1}^{r} a_i \left(\sum_{j=1}^{k} p_j^2\right) \qquad [2.4.4]$$

pues son iguales salvo una constante y dan lugar a la misma ambigüe dad.

En [2.4.4] la función convexa C que genera la ambigüedad es:

$$c(p) = \sum_{i=1}^{k} p_i^2$$

que es la energía informacional y que hemos visto que proporciona la misma ambigüedad que la distancia euclídea.

# CAPITULO III

3.1. Algoritmos basados en la segmentación

Entre las técnicas más adecuadas para la predicción de una variable dependiente y en función de un conjunto de variables explicativas, cuando estas son categóricas y la variable dependiente puede ser de intervalo, ordinal o categórica, estan las técnicas de segmentación. En este sentido puede consultarse a Sonquist y Morgan (1964), Messenger y Mandell (1972) y Sánchez (1980).

Todo proceso de segmentación se basa en la siguiente idea:

Se toma una variable explicativa y se divide la población o conjunto de individuos observados en $k$ subconjuntos -normalmente se suele tomar $k=2$- de forma que se maximice el objetivo previamente fijado. Este proceso se realiza con cada una de las variables explicativas, tomando aquella que haya maximizado la función objetivo. En cada uno de los $k$ segmentos en que ha quedado particionado el conjunto de individuos observados se procede de forma similar con el conjunto de variables explicativas que no han participado en dicha partición. El proceso se repite siempre que los conjuntos a particionar tengan un número de elementos superior a un valor prefijado.

Esta idea es la que sirve de base a los algoritmos que a continuación se desarrollan para obtener la solución al problema de discriminación tal y como lo hemos planteado.

Definición 3.1.1.- $\forall D \subset \prod_{i=1}^{p} K\xi_i$, diremos que $D$ es un segmento cartesiano cuando $D$ puede expresarse como:

$$D = A_1\ \xi_1 \times A_2\ \xi_2 \times \ldots \times A_p\ \xi_p \qquad [3.1.1]$$

donde:

$$A_i\ \xi_i \subset K\xi_i, \qquad i=1,\ldots,p$$

A toda solución S de (P,S) que esté formada por segmentos cartesianos, la llamaremos solución cartesiana.

Dado un segmento cartesiano $D = A_1\ \xi_1 \times \ldots \times A_p\ \xi_p$ diremos que D es g-$\ell$-segmentable cuando $\exists\ \xi_j \in E$ / $\exists$ una partición propia de $A_j\ \xi_j$; $(A_{j_1}\ \xi_j, \ldots, A_{j_\ell}\ \xi_j)$ que cumple:

$$f(D \cap (K\ \xi_1 \times \ldots \times A_{j_r}\ \xi_j \times \ldots \times K\xi_p)) \geq g \quad \forall r=1,\ldots,\ell \qquad [3.1.2]$$

donde f es la función peso.

Es inmediato comprobar que si D es un segmento g-$\ell$--segmentable entonces es $g_1$-$\ell_1$-segmentable $\forall \ell_1 \leq \ell$ y $g_1 \leq g$.

En general diremos que un segmento es g-segmentable cuando $\exists$ algún $\ell \in N$ tal que sea g-$\ell$-segmentable y de aquí en adelante nos referiremos indistintamente a la g-segmentabilidad como segmentabilidad y viceversa.

El concepto de segmentabilidad para segmentos nos lleva de forma natural a la idea de segmentación de soluciones y diremos que una solución cartesiana S puede segmentarse ó es segmentable cuando tenga algún segmento segmentable.

Sea S una solución cartesiana, designamos por T(S):

T(S) = {S'/S' se obtiene de S segmentando un solo segmento de S}.

Evidentemente se cumple:

$$S \text{ es segmentable } \Longleftrightarrow T(S) \neq \emptyset.$$

<u>Definición 1.3.2</u>.- S, solución cartesiana es terminal para la función bondad B si se cumple que $T(S) = \emptyset$ ó $B(S') < B(S)$ $\forall S' \in T(S)$.

La definición anterior caracteriza máximos locales para las funciones bondad en el sentido de que habremos llegado a un máximo local para la función B a través de la segmentación cuando no se pueda mejorar esta mediante la segmentación.

Habiendo precisado el concepto de máximo local para las funciones bondad, el objetivo de aquí en adelante será encontrar el modo de llegar a las soluciones terminales que me dan dichos máximos locales.

La forma más fácil será caracterizando los segmentos de una solución terminal.

Sea S, solución cartesiana y $S = (D_1, \ldots, D_\ell)$, designamos por:

$$T_{D_i}(S) = \{S' \in T(S),\ S' \text{ se obtiene de } S \text{ segmentado } D_i\}$$

se cumple:

$$\bigcup_{i=1}^{\ell} T_{D_i}(S) = T(S)$$

$$T_{D_i}(S) \cap T_{D_j}(S) = \emptyset, \quad i \neq j$$

Introducimos la noción de segmento terminal en una solución S mediante la siguiente definición:

<u>Definición 3.1.3</u>.- Sea $S$, solución cartesiana y $S = (D_1,\dots,D_\ell)$, diremos que $D_i$ es un segmento terminal de $S$ para la función bondad $B$ si se cumple que $T_{D_i}(S) = \emptyset$ ó $B(S') < B(S)$ $\forall S' \in T_{D_i}(S)$.

De la definición anterior se deduce de forma inmediata que si una solución cartesiana $S$ tiene todos sus segmentos terminales dicha solución es terminal. Es claro por tanto que si se encuentra un método que determine soluciones cuyos segmentos sean todos terminales este método encontrará soluciones que sean máximos locales.

Distinguiremos dos situaciones dependiendo de la función bondad.

## 3.2. Algoritmos para funciones de tipo B1.

En el caso de que las funciones bondad sean de tipo B1, se cumplen las siguientes propiedades:

P.1 $S$, solución cartesiana es terminal para $B \iff T(S) = \emptyset$

<u>Demostración</u>: Supongamos que $S$ sea terminal para $B$ y $T(S) \neq \emptyset$ $\Longrightarrow$ $\forall S' \in T(S)$ es $B(S') < B(S)$.

Si $S' \in T(S) \Longrightarrow S \lneq S'$ y por ser $B$ de tipo B1 $\Longrightarrow$ $\Longrightarrow B(S) \leq B(S')$ llegando a una contradicción.

Análogamente demostraríamos:

P.2 Sea $S$ solución cartesiana y $S = (D_1,\dots,D_\ell)$, se tiene que $D_i$ es un segmento terminal de $S$ para $B$ si y sólo si $T_{D_i}(S) = \emptyset$.

La consecuencia más importante de estas propiedades es que podemos caracterizar a los segmentos terminales independientemente de la solución a la que pertenezcan. En efecto:

Sea D un segmento terminal de S, según la propiedad P2 se tiene que $T_D(S) = \emptyset$ y esto quiere decir que no va a existir ninguna solución que se deduzca de S segmentando D, es decir, que D no es segmentable. Luego los segmentos terminales de una solución S' cuando las funciones son de tipo Bi coinciden con los segmentos no segmentables, en cuya definición no se hace referencia a la solución.

Este resultado es de suma importancia a la hora de buscar las soluciones terminales pues bastará con encontrar segmentos que no se puedan segmentar y que formen una partición de $\Pi K\xi_i$. Y es debido a este resultado por lo que el siguiente algoritmo llega a una solución terminal.

Utilizaremos la siguiente notación:

Sea $S = (D_1, D_2, \ldots, D_\ell)$, una solución cartesiana, por [3.1.1] cada segmento de S puede expresarse por:

$$D_i = A_1^i \, \xi_1 \times \ldots \times A_p^i \, \xi_p, \qquad i=1,\ldots,\ell$$

Por analogía a determinados lenguajes de programación representaremos a S por:

$$S(i,j) = A_j^i \, \xi_j, \qquad i=1,\ldots,\ell \quad j=1,\ldots,p$$

Si designamos por $P_t(S(i,j))$ al conjunto de particiones de $S(i,j)$, $\forall A \in P_t(S(i,j))$ t.q. $A = (A_1, A_2, \ldots, A_{|A|})$ denotamos por

$S_A \in T_{D_i}(S)$ a la solución obtenida de segmentar $S$ por $A$ y la definimos de la siguiente forma:

$$S_A(\ell_1,\ell_2) = \begin{cases} S(\ell_1,\ell_2) & \text{si} \quad \ell_1 < i \quad y \quad \ell_2 = 1,\dots,p \\ S(\ell_1+1,\ell_2) & \text{si} \quad i < \ell_1 < \ell \quad y \quad \ell_2 = 1,\dots,p \\ A_{\ell_1-\ell+1} & \text{si} \quad \ell_1 = \ell,\dots,\ell+|A|-1 \quad y \quad \ell_2 = j \\ S(i,\ell_2) & \text{si} \quad \ell_1 = \ell,\dots,\ell+|A|-1 \quad y \quad \ell_2 = 1,\dots \\ & \qquad \dots,j-1,\ j+1,\dots,p \end{cases}$$

Para ver si la solución $S_A$ pertenece al conjunto de soluciones g-admisibles será suficiente con comprobar que el segmento i-ésimo es segmentable a través de $A$ es decir:

$$S_A \in A_g(P,S) \iff f(\prod_{\ell_2=1}^{p} S(\ell_1,\ell_2)) \geq g \quad \forall \ell_1 = \ell,\dots,\ell+|A|-1$$

Sea

$$P_{tg}(S(i,j)) = \{A \in P_t(S(i,j)) / S_A \in A_g(P,S)\}$$

Si $P_{tg}(S(i,j)) \neq \emptyset$ designamos por $A(i,j)$ la partición de $S(i,j)$ que cumple con:

$$A(i,j) \in P_{tg}(S(i,j)) \ / \ B(S_{A(i,j)}) = \max_{A \in P_{tg}(S(i,j))} B(S_A)$$

El algoritmo que se propone, en base a la anterior notación queda en los siguientes términos:

<u>Algoritmo 1</u>.

<u>Paso 0</u> (De inicialización)

$T(S) = 1$

$I = \{1\}$

$J(i) = \{1,2,\dots,p\}, \quad \forall i \in I$

$S(i,j) = K\xi_j, \quad i \leq T(S) \quad y \quad j=1,2,\dots,p$

$F = 0$

$IC = \emptyset$

Paso 1 se hace $\forall i \in I, \quad \forall j \in J(i)$

Calculamos $P_{tg}(S(i,j))$

Si: $P_{tg}(S(i,j)) = \emptyset$ entonces $J(i) = J(i) - \{j\}$

$P_{tg}(S(i,j)) \neq \emptyset$ se calcula $S_{A(i,j)}$

Paso 2 Se hace $\forall i \in I$

Si $J(i) = \emptyset$ entonces $I = I - \{i\}$

$F = F + 1$

$$D_F = \prod_{j=1}^{p} S(i,j)$$

Paso 3

Si: $I = \emptyset$, entonces parar, $(D_1,\dots,D_F)$ es solución terminal

$I \neq \emptyset$ ir a Paso 4

Paso 4

Se busca $(i_o,j_o)$ con $i_o \in I$ y $j_o \in J(i_o)$ que cumpla:

$$B(S_{A(i_o,j_o)}) = \underset{i \in I}{\mathrm{Max}} \ \underset{j \in J(i)}{\mathrm{Max}} \ B(S_{A(i,j)})$$

Paso 5 Se hace $\forall i \in I$

Si $i \in I$ y $i < i_o$ entonces $IC = IC \cup \{i\}$

Si $i \in I$ y $i > i_o$ entonces $IC = IC \cup \{i-1\}$ y $J(i-1) = J(i)$

Paso 6

$I = IC \cup \{T(S),\ T(S)+1, \ldots, T(S)+|A(i_o,j_o)|-1\}$

$J(i) = \{1,2,\ldots,p\} \quad \forall i \in \{T(S), T(S)+1, \ldots, T(S)+|A(i_o,j_o)|-1\}$

$S = S_{A(i_o,j_o)}$

$IC = \emptyset$

$T(S) - T(S) + |A(i_o,j_o)| - 1$

Ir a paso 1

El anterior algoritmo, aunque basado en los algoritmos de segmentación, presenta dos variaciones sobre éstos, lo que supone una generalización en su concepción, éstas variantes viene dadas por:

1ª) El número de segmentos en que en cada etapa se particiona el conjunto de items observados, no es fijado a priori, si no que varía dependiendo de la partición de la variable que da lugar a dicha segmentación. (En el algoritmo anterior viene dado por $|A(i_o,j_o)|$, calculado en el paso 4).

2ª) Cuando una variable ha provocado la segmentación en una etapa, dicha variable puede generar una nueva segmentación, a diferencia de lo que ocurre en los algoritmos de segmentación. (En el algoritmo anterior, esto se debe al hecho de ser $J(i) = \{1,2,\ldots,p\}$ $\forall i \in \{T(s),\ldots,T(s) + |A(i_o,j_o)| - 1\}$).

3.3. Algoritmos para funciones bondad separables.

En el párrafo anterior se ha visto que una de las ventajas desde el punto de vista algorítmico que presentan las funciones Bl se basa en el hecho de que la caracterización de los segmentos terminales se haga independientemente de la solución a la que pertenezcan. En este sentido vamos a demostrar que para el conjunto de funciones bondad que sean separables se obtiene un resultado análogo al obtenido para las funciones de tipo Bl.

Definición 3.3.1.-Diremos que una función bondad $B$ es separable cuando $B$ puede expresarse:

$$B(S) = \sum_{i=1}^{r} Z(D_i), \quad S \in A_g(P,S)/S = (D_1,\dots,D_r)$$

donde:

$$Z : G \longrightarrow R^+ \cup \{0\}$$

y

$$G = \{D/D \subset \Pi K\xi_i, \quad f(D) \geq g\}.$$

La condición de separabilidad parece bastante natural desde un punto de vista práctico y puede comprobarse que las funciones bondad consideradas en los apartados 1.5 y 2.4 cumplen con dicha condición. Para estas funciones se cumple el siguiente resultado:

Teorema 3.3.1.- Si $D$ es un segmento terminal de $S$ para $B$ y $B$ es una función separable entonces $D$ es segmento terminal para $B$ de cualquier solución $S'$ que lo tenga como segmento.

<u>Demostración</u>: Sean $S = (D,D_1,\dots,D_r)$ y $S' = (D,D_1',\dots,D_{r'}')$ se tiene:

$$T_D(S) = \emptyset \iff T_D(S') = \emptyset$$

y si $T_D(S) \neq \emptyset$, entonces:

$$\forall S'' \in T_D(S) \text{ es } B(S'') < B(S) \iff \sum_{h=1}^{r} Z(D_h) + \sum_{h=1}^{\ell} Z(D^h) <$$

$$< Z(D) + \sum_{h=1}^{r} Z(D_h) \iff \sum_{h=1}^{\ell} Z(D^h) < Z(D) \iff \sum_{h=1}^{\ell} Z(D^h) +$$

$$+ \sum_{h=1}^{r'} +Z(D_h') < Z(D) + \sum_{h=1}^{r'} Z(D_h') \iff B(S''') < B(S)$$

donde:

$$S'' = (D^1,D^2,\dots,D^{\ell},D_1,\dots,D_r) \in T_D(S)$$

$$S''' = (D^1,D^2,\dots,D^{\ell},D_1',\dots,D_{r'}') \in T_D(S').$$

#

Análogamente a como definimos el conjunto $T_D(S)$, definimos el conjunto $T_D^j(S)$ de la siguiente forma:

Sea $S$, solución cartesiana y $S = (D_1,\dots,D_{\ell})$, $\forall j=1,2,\dots,p$ es:

$$T_{D_i}^j(S) = \{S' \in T_D(S) \ / \ S' \text{ se obtiene de } S \text{ segmentando } D \text{ a través de la variable } \xi_j\}$$

Se cumple:

$$\sum_{j=1}^{p} T_{D_i}^j(S) = T_D(S)$$

$$T_{D_i}^{j_1}(S) \cap T_{D_i}^{j_2}(S) = \emptyset \quad \text{si} \quad j_1 \neq j_2.$$

Introducimos la noción de variable que es terminal para un segmento $D$ de una solución $S$ mediante la siguiente definición:

Definición 3.3.1.- Sea S, solución cartesiana y D un segmento de S, diremos que la variable $\xi_j$ es terminal para D en S para la función bondad B si se cumple que $T_D^j(S) = \emptyset$ ó $B(S') < B(S)$ $\forall S' \in T_D^j(S)$.

De la definición anterior se deduce de forma inmediata que un segmento D es terminal en S respecto de B si y sólo si todas las variables son terminales para D en S respecto de B.

Análogamente al teorema 3.3.1 se demuestra el siguiente teorema:

Teorema 3.3.2.- Si $\xi_j$ es terminal para D en S respecto de B y B es una función separable, entonces $\xi_j$ es terminal para D en S' respecto de B para cualquier otra solución S', cartesiana que tenga a D como segmento.

Los resultados anteriores nos permiten desarrollar algoritmos que nos determinen soluciones terminales a través de la localización sucesiva de segmentos terminales y la posibilidad de no tener que probar variables que se hubiesen rechazado en la localización de estos.

Lo anteriormente dicho justifica el paso 1 del siguiente algoritmo que es idéntico al algoritmo 1 salvo en el paso 1.

Algoritmo 2

Se hace igual que en el algoritmo 1, quedando el 1° paso en la siguiente forma:

Paso 1. Se hace $\forall i \in I$, $\forall j \in J(i)$

Calculamos $P_{tg}(S(i,j))$

Si $P_{tg}(S(i,j)) = \emptyset$ entonces $J(i) = J(i) - \{j\}$

$P_{tg}(S(i,j)) \neq 0$ se calcula $S_{A(i,j)}$

Si $B(S_{A(i,j)}) < B(S)$ entonces

$J(i) = J(i) - \{j\}$

3.4. Algoritmo general

Describimos aquí, un algoritmo que nos permita calcular soluciones terminales para cualquier función bondad. Este algoritmo es idéntico al algoritmo uno salvo que el Paso 4 toma la siguiente forma:

Paso 4

Se busca $(i_o,j_o)$ con $i_o \in I$ y $j_o \in J(i_o)$ que cumpla:

$$B(S_{A(i_o,j_o)}) = \max_{i\in I} \max_{j\in J(i)} B(S_{A(i,j)})$$

Si $B(S_{A(i_o,j_o)}) < B(S)$ Parar, $S$ es solución terminal

$B(S_{A(i_o,j_o)}) \geq B(S)$ ir a paso 5

3.5. Generalización de la solución y algoritmo.

En el apartado 3.3 introducimos la noción de segmento terminal, este venía asociado a una solución y dependía de la función bondad que se tomaba. El hecho de que para las funciones bondad separables estos fuesen independientes de la solución sugiere, el determinar estos sin necesidad de ir asociados a solución alguna a diferencia de lo que se hace en el algoritmo dos. Determinar segmentos terminales sin tener en cuenta la solución a la que pertenecen conduce a la introducción de una función valor de segmento ya que la terminalidad de un segmento depende de la función bondad y esta solo estaba definida para soluciones, lo que obligaba ir arrastrando una solución cuando determinábamos los segmentos terminales. La función valor nos permitirá desligarnos de las soluciones para la determinación de dichos segmentos.

Supuesto que tenemos determinada una colección de segmentos terminales para una situación (P,S) y que estos, no vengan asociados a ninguna solución, cuando estos forman un recubrimiento de $\Pi K \xi_i$, esta colección nos define un tipo de solución más general que las consideradas hasta ahora, pues admite la posibilidad de que dichos segmentos no sean disjuntos. Esta propiedad de dichas soluciones generales es la que justifica su determinación pues es muchas veces una visión más realista del problema dado que los solapamientos entre los segmentos que definen dicha solución van a dar información sobre la ambigüedad inherente al problema.

Considerar este tipo de soluciones además de acarrear problemas de tipo algorítmico a la hora de determinar los segmentos que la definen, nos lleva a la generalización de las funciones bondad hasta ahora consideradas, para que englobando a dichas soluciones podamos compararlas, y al problema de generación de clasificadores a partir de dichas soluciones, que tengan en cuenta la ambigüedad dada por los solapamientos.

El objetivo que nos hemos propuesto en este apartado es dar una solución a dichos problemas.

### 3.5.1. Función valor para segmentos.

Sea $\beta$ una función del simplex $\Delta \subset R^k$ en los reales positivos

$$\beta : \Delta \longrightarrow R^+ \cup \{0\}$$

Definimos valor de un segmento D, asociado a la función $\beta$, y lo designamos por $U(D)$ a:

$$U : G \longrightarrow R^+ \cup \{0\}$$

$$D \longrightarrow U(D) = \beta(p_D)$$

donde G es como en la definición 3.3.1 el conjunto de segmentos con peso mayor o igual a g.

La definición anterior nos da una idea precisa -dependiendo del tipo de función $\beta$ elegida- de la importancia que tiene un segmento a la hora de estimar el valor de $\eta$ a través de dicho segmento.

Un caso particular de funciones valor que podemos considerar es el que viene generado por la ambigüedad de los segmentos visto en el capítulo dos, en este caso la función U viene definida por:

$$U(D) = M - A_D(H), \qquad D \in G$$

Otra especificación de U, y con relación a los resultados obtenidos en el capítulo primero sobre las funciones bondad B1, es la proporcionada cuando la función $\beta$ asociada es una función convexa en $\Delta$.

A toda función valor U podemos asociarla una función bondad $B_U$ definida por:

$$B_U(S) = \frac{1}{f(\Pi K\xi_i)} \sum_{i=1}^{\ell} f(D_i)U(D_i), \quad S = (D_1,\dots,D_\ell) \qquad [3.5.1]$$

Se cumple que si, la función $\psi$ que genera la función $S$ viene definida por [1.2.1], y las funciones pesos por [1.2.12] la función $B_U$ tiene el mismo comportamiento que las funciones bondad de tipo B1 caracterizadas en el teorema 1.4.1.1 cuando la función $\beta$ asociada a $U$ es una función convexa en $\Delta$, ya que ambas coinciden salvo la constante de proporcionalidad $\frac{1}{f(\Pi K\xi_i)}$. Con lo que conectamos este tipo de funciones con las estudiadas en los capítulos anteriores.

### 3.5.2. Generación de funciones valor a través de funciones bondad.

Dada la situación $(P,S)$ con $P = (I,E,\eta)$, supondremos en lo que sigue que la función peso $f$ asociada a $(P,S)$ viene definidad por [1.2.12] y que la función $S$ viene generada a través de la función $\psi$ definida en [1.2.1].

Para todo subconjunto de items $I' \subset I$ designamos por $(P_{I'},S_{I'})$ al subproblema definido por: $P_{I'} = (I',E,\eta)$ y $S_{I'}$ definida por [1.2.3] a través de $\psi$.

La función peso asociada a $(P_{I'},S_{I'})$ la designamos por analogía con $f_{I'}$ y viene dada por [1.2.12].

Análogamente a la definición de $A_g(P,S)$, designamos por $A_g(P_{I'},S_{I'})$ al conjunto de soluciones admisibles para $(P_I,S_I)$ con la función peso $f_I$.

Supondremos que dada una función bondad B sobre $A_g(P,S)$, podemos restringirla a $A_g(P_{I'},S_{I'})$, designando por $B(I',S)$ el valor de la función bondad $B$ para $S \in A_g(P_{I'},S_{I'})$. Análogamente para la función $U$, utilizamos la notación $U(I',D)$ para precisar sobre que subproblema consideramos el valor del segmento.

Dada $(P,S)$, respresentamos por $S_P$ a la solución de $A_g(P,S)$ que contiene un único segmento, es decir aquella que estima $\eta$ a partir del conjunto de items sin basarse en la información proporcionada por $E$.

Bajo estas consideraciones, demostramos el siguiente resultado para las funciones $B_U$ definidas en [3.5.1].

Teorema 3.5.1.- Toda función bondad $B_U$, definida como en [3.5.1] puede expresarse como:

$$B_U(I,S) = \sum_{i=1}^{r} \frac{f(D_i)}{f(\Pi K\xi_i)} \; B_U(\hat{I}(D_i),\, S_{P_{\hat{I}(D_i)}}), \quad S=(D_1,\ldots,D_r) \qquad [3.5.2]$$

Demostración.- Será suficiente con demostrar que:

$$B_U(\hat{I}(D), S_{P_{\hat{I}(D)}}) = U(I,D) \quad \forall D \subset \Pi K\xi_i \quad \text{y} \quad f(D) \geq g$$

En efecto:

$$B_U(\hat{I}(D),\, S_{P_{\hat{I}(D)}}) = \frac{f_{\hat{I}(D)}\,(\Pi K\xi_i)}{f_{\hat{I}(D)}\,(\Pi K\xi_i)}\; U(\hat{I}(D),\, \Pi K\xi_i)$$

y como se cumple:

$$U(\hat{I}(D),\, \Pi K\xi_i) = U(I,D)$$

pues $D$ de $(P,S)$ y $\Pi K\xi_i$ de $(P_{\hat{I}(D)},\, \hat{I}(D))$ se aplican en el mismo punto de $\Delta$ ya que $S$ y $S_{\hat{I}(D)}$ vienen definidas a partir de [1.2.3]. Teniendo por tanto que dan el mismo valor para la función $\beta$.

Se tiene:

$$B_U(\hat{I}(D),\ S_{P_{\hat{I}(D)}}) = U(I,D)$$

Este resultado nos permite generar funciones valor de segmentos a través de cualquier función bondad B que tengamos y que designaremos por $U_B$.

$$U_B(D) = B(\hat{I}(D),\ S_{P_{\hat{I}(D)}}) \quad D \subset \Pi K\xi_i \quad y \quad f(D) \geq g \qquad [3.5.3]$$

Sean $\mathcal{B}$ el conjunto de todas las funciones bondad y $\mathcal{U}$ el conjunto de todas las funciones valor de segmentos las relacciones [3.5.1] y [3.5.3] nos conducen a definir una aplicación suprayectiva $\rho$ entre $\mathcal{B}$ y $\mathcal{U}$ definida por:

$$\rho : \mathcal{B} \longrightarrow \mathcal{U}$$

$$\rho(B)(D) = B(\hat{I}(D),\ S_{P_{\hat{I}(D)}}) \quad \forall B \in \mathcal{B} \quad y \quad D \subset \Pi K\xi_i$$

Evidentemente la $\rho$ así definida es aplicación.

Es suprayectiva:

$$\forall U \in \mathcal{U},\ \exists B \in \mathcal{B} \ / \ \rho(B) = U$$

En efecto:

Sea $U \in \mathcal{U}$, definimos $B_U(S)$ como en [3.5.1], y por el teorema 3.5.1 se cumple que $\rho(B_U) = U$.

Esta aplicación nos establece en $\mathcal{B}$ una relación de equivalencia, que relacciona dos funciones bondad cuando su imagen a través de $\rho$ es la misma y puede expresarse por:

$$B_1, B_2 \in \mathcal{B}, \quad B_1 \, r \, B_2 \iff B_1(\hat{I}(D), S_{P_{\hat{I}(D)}}) = B_2(\hat{I}(D), S_{P_{ID)}})$$

$$\forall D \subset \Pi K \xi_i \quad \text{t.q.} \quad f_I(D) \geq g$$

Tomaremos como representante de la clase la definida por [3.5.2].

La aplicación más importante del resultado anterior es que dada una situación (P,S) -a esta la podemos dotar de una bondad definida en $A_g(P,S)$ ó de una función valor definida en G- si suponemos que tenemos definida una función valor y queremos atacar el problema a través de funciones bondades bastará con construir una bondad a partir de la función valor definida por [3.5.1]. De forma similar haríamos si partimos de una función valor. Se tiene además la posibilidad de contrastar la coherencia del punto de partida cuando este consiste en una función valor y una función bondad, podemos decir que son coherentes cuando la función bondad deducida a partir de la función valor pertenece a la misma clase de equivalencia que la función bondad de partida.

3.5.3. Segmentos terminales.

Análogamente a lo hecho en el apartado 3.5.2, denotaremos por $T_D(I',S)$ el conjunto $T_D(S)$ para el subproblema $(P_{I'}, S_{I'})$. Para (P,S), utilizamos indistintamente $T_D(I,S)$ ó $T_D(S)$.

El siguiente resultado válido para funciones bondad que puedan expresarse por [3.5.3], y en particular por el teorema 3.5.1 válido para las funciones bondad definidas a partir de funciones valor mediante [3.5.1], nos va a permitir caracterizar los segmentos terminales para funciones valor.

Teorema 3.5.2.- Para cualquier función bondad $B \in \mathcal{B}$ de $(P,S)$, que pueda expresarse como en [3.5.3], se cumple que $D \in G$ es segmento terminal para $B \Longleftrightarrow S_{P_{\hat{I}(D)}}$ es solución terminal para $B$ definida en $A_g(P_{\hat{I}(D)}, S_{\hat{I}(D)})$.

Demostración.- Por ser $B$ separable y aplicando el resultado del teorema 3.3.1 puede hablarse de terminalidad de $D$ independientemente de la solución a la que pertenezca, en particular $D$ es terminal para $(D, \Pi K\xi_i - D)$.

Sea $S = (D, \Pi K\xi_i - D)$, se tiene:

$D$ es terminal para $S$ a través de $B \Leftrightarrow T_D(S) = \emptyset$ ó $B(S') < B(S) \quad \forall S' \in T_D(S)$.

Se cumple:

$$T_D(S) = \emptyset \quad \Longleftrightarrow \quad T(\hat{I}(D),\ S_{P_{\hat{I}(D)}}) = \emptyset$$

Si $T_D(S) \neq \emptyset$, $\forall S' \in T_D(S)$ $\exists$ $S'_{\hat{I}(D)} \in T(\hat{I}(D), S_{P_{\hat{I}(D)}})$ que se obtiene de $S_{P_{\hat{I}(D)}}$ segmentando $\Pi K\xi_i$ a través de la misma variable y por la misma partición con la que obteníamos $S'$ a través de $S$. Análogamente $\forall S'_{\hat{I}(D)} \in T(\hat{I}(D), S_{P_{\hat{I}(D)}})$ $\exists\, S' \in T_D(S)$ que se obtiene segmentando $S$ a través de $D$ por la misma variable y por la misma partición con la que obteníamos $S'_{\hat{I}(D)}$. Con este resultado establecemos una correspondencia biunívoca entre los conjuntos $T_D(S)$ y $T(\hat{I}(D), S_{P_{\hat{I}(D)}})$.

Para demostrar el teorema será suficiente con ver:

Si $T_D(S) \neq \emptyset$ y $D$ es terminal a través de $B$ entonces

$\forall S'_{\hat{I}(D)} \in T(\hat{I}(D),\ S_{P_{\hat{I}(D)}})$ se cumple que $B(\hat{I}(D),\ S'_{\hat{I}(D)}) <$ $< B(\hat{I}(D),\ S_{P_{\hat{I}(D)}})$.

En efecto:

$\forall S'_{\hat{I}(D)} \in T(\hat{I}(D),\ S_{P_{\hat{I}(D)}})\ \exists\ S' \in T_D(S)$ obtenido con la misma partición de la misma variable que la que obtiene $S'_{\hat{I}(D)}$. Sean $(D_1,\ldots,D_\ell)$ los segmentos que resultan de segmentar $D$, se tiene:

$$B(S) = \frac{f(\Pi K\xi_i - D)}{f(\Pi K\xi_i)}\, B(\hat{I}(\Pi K\xi_i - D),\quad S_{P_{\hat{I}(\Pi K\xi_i - D)}}) +$$

$$+\ \frac{f(D)}{f(\Pi K\xi_i)}\, B(\hat{I}(D),\quad S_{P_{\hat{I}(D)}})$$

$$B(S') = \frac{f(\Pi K\xi_i - D)}{f(\Pi K\xi_i)}\, B(\hat{I}(\Pi K\xi_i - D),\quad S_{P_{\hat{I}(\Pi K\xi_i - D)}}) +$$

$$+\ \sum_{i=1}^{\ell} \frac{f(D_i)}{f(\Pi K\xi_i)}\, B(\hat{I}(D_i),\ S_{P_{\hat{I}(D_i)}})$$

Por ser $D$ terminal de $S$ para $B$ se tiene que $B(S') <$ $< B(S)$ luego:

$$\sum_{i=1}^{\ell} \frac{f(D_i)}{f(\Pi K\xi_i)}\, B(\hat{I}(D_i),\ S_{P_{\hat{I}(D_i)}}) \qquad \frac{f(D)}{f(\Pi K\xi_i)}\, B(\hat{I}(D),\quad S_{P_{\hat{I}(D)}})$$

$$\sum_{i=1}^{\ell} \frac{f(D_i)}{f(D)}\, B(\hat{I}(D_i),\ S_{P_{\hat{I}(D_i)}}) < B(\hat{I}(D),\ S_{P_{\hat{I}(D)}})$$

y como $f(D) = f_{I(D)}\ (\Pi K\xi_i)$

$f(D_i) = f_{\hat{I}(D)}\ (D_i)$ pues $D_i \subset\subset D$.

Tenemos:

$$B(\hat{I}(D),\ S'_{\hat{I}(D)}) < B(I(D),\quad S_{P_{\hat{I}(D)}}).$$

Una consecuencia importante del anterior teorema es que en el caso particular de que las funciones separables puedan expresarse como en [3.5.3] la terminalidad de cualquier segmento D se caracteriza por la terminalidad de $S_{P_{\hat{I}(D)}}$ . Con lo que podemos desligarle de cualquier solución.

Asi mismo, este resultado nos lleva de forma natural a definir lo que entendemos por segmento terminal a través de una función valor U, sin tener que asociarlo a ninguna solución, consiguiendo uno de los objetivos propuestos.

<u>Definición 3.5.1</u>.- Diremos que $D \in G$ es terminal a través de la función U cuando $S_{P_{\hat{I}(D)}}$ es terminal para la función bondad $B_U$ relativa a $(P_{\hat{I}(D)}, S_{\hat{I}(D)})$ y $B_U$ definida por [3.5.1].

Consecuencia inmediata de esta definición y del teorema 3.5.2 es que cuando B es expresable como en [3.5.3] un segmento D es terminal para B si y sólo si D es terminal para la función valor $\rho(B)$.

3.5.4. Generalización de la solución.

Si suponemos que en (P,S) tenemos definida una función valor de segmento U, por el párrafo anterior podemos caracterizar los segmentos terminales para U; esto nos induce a considerar soluciones que estan basadas en segmentos terminales para U, y al no tener que ir asociados estos segmentos a particiones, no tienen por que formar ellos a su vez una partición, generalizando el concepto de solución de (P,S) a considerar que estos pueden basarse en recu brimientos de $\Pi K\xi_i$ y no unicamente en particiones.

Definición 3.5.2.- Diremos que $\mathcal{D}$ determina una solución de (P,S) si, $\mathcal{D}$ es una familia de subconjuntos de $\Pi K\xi_i$ cumpliendo:

1) $\bigcup\limits_{D \in \mathcal{D}} D = \Pi K\xi_i$

2) $\nexists\ D_1, D_2 \in \mathcal{D}\ /\ D_1 \cap D_2 \neq \emptyset$ y $D_1 \subset D_2$ ó $D_2 \subset D_1$.

La definición anterior generaliza el concepto de solución en el sentido en que $\mathcal{D}$ no determina de forma única la solución ya que si existen elementos de la familia con intersección no vacía los puntos de dichas intersecciones no se aplican en un único punto de $\Delta$, dejando cierta difusidad en la determinación de esta, razón por la cual esta forma de generar la solución responde de manera más realista a las situaciones prácticas.

Una forma de determinar una solución de (P,S) a partir de una familia $\mathcal{D}$ cumpliendo con la definición 3.5.2 viene dada por $S_{\mathcal{D}}$:

$$S_{\mathcal{D}}(x) = \sum_{j=1}^{r} \frac{f(D_i)}{\sum\limits_{j=1}^{r} f(D_j)} P_{D_{ij}} \qquad [3.5.4]$$

donde:

$$x \in \prod_{i=1}^{p} K\xi_i \quad \text{y} \quad x \in \bigcup_{j=1}^{r} D_{ij} \quad \text{con} \quad D_{i_1}, \ldots, D_{i_r} \in \mathcal{D} \quad \text{y}$$

$$P_{D_{ij}} \quad \text{cumple con } [1.2.8]$$

Evidentemente esta forma de determinar la solución a partir de $\mathcal{D}$ no es la única posible, pero cumple con una condición desea ble en toda determinación pues cuando $\mathcal{D}$ es una partición de $\Pi K\xi_i$, $S_{\mathcal{D}}$ coincide con las soluciones hasta ahora consideradas.

Otra forma de expresar esta generalización, es que la $S_{\mathcal{D}}$ determinada por [3.5.4], puede tener una partición asociada que no sea admisible aunque la familia $\mathcal{D}$ esté formada por segmentos que pertenezcan a G que es la condición que impondremos a la hora de buscar una familia $\mathcal{D}$ g-admisible.

Análogamente, si queremos dar una bondad a la solución derivada a partir de una familia $\mathcal{D}$, consideraremos aquella que cuando $\mathcal{D}$ sea una partición de $\Pi K\xi_i$ con todos sus elementos de G y por tanto solución g-admisible, coincida con la función bondad $B_U$ definida en [3.5.1]. Esta función bondad para $S_{\mathcal{D}}$ viene dada por:

$$B_U^*(\mathcal{D}) = \frac{1}{\sum_{i=1}^{r} f(D_i)} \sum_{i=1}^{r} f(D_i)U(D_i)$$

y evidentemente cumple con [3.5.1] cuando $\mathcal{D}$ es una partición g-admisible de $\Pi K\xi_i$.

A la hora de determinar una familia $\mathcal{D}$ que sea máximo local para U, nos ceñimos a buscarla entre las familias formadas por segmentos cartesianos de G. Esta por el párrafo anterior, vendrá dada por aquella que tenga todos sus segmentos terminales para U y que por el teorema 3.5.2, cuando $\mathcal{D}$ sea una partición de $\Pi K\xi_i$ y por tanto una solución de $A_g(P,S)$ es a su vez un máximo local para la función bondad $B_U$ definida en [3.5.1].

3.5.5. Algoritmo para determinar segmentos terminales para una función valor.

En el desarrollo del algoritmo que se describe en este apartado necesitamos introducir el concepto de segmento cartesiano U-terminal.

Sea $D \in G$ y $D$ segmento cartesiano, designamos por $t(D)$

$$t(D) = \{D' \in t(D)/D' \text{ se obtiene de } D \text{ segmentando } D\}$$

Diremos que un segmento $D$ es U-terminal cuando $t(D) = \emptyset$ ó bien $U(D') < U(D)$ $\forall D' \in t(D)$.

Es fácil comprobar el siguiente resultado:

"Si $D$ es U-terminal entonces $D$ es terminal para $U$".

Como consecuencia de lo anterior si obtengo una familia de segmentos U-terminales, esta, estará formada por segmentos terminales para $U$ y por tanto será un máximo local.

El siguiente algoritmo, encuentra una familia de segmentos -terminales, y utiliza la siguiente notación:

Designaremos por $D^* = \{(i_1,j_1),\dots(i\ ,j_\ell)\}$ al segmento cartesiano $D = A_1\xi_1 \times\dots\times A_p\xi_p$ tal que $A_i\xi_i = K\xi_i$ $\forall i \notin \{i_1,\dots,i_\ell\}$ y $A_{i_r}\xi_{i_r} = \{j_r\}$ $\forall i_r \in \{i_1,\dots,i_\ell\}$ y por $U(D^*)$ a la valor de la función $U$ para el segmento $D$.

Dado $(D_1,\dots,D_\ell)$, designaremos por $(D_1,\dots,D_\ell)^C$ al segmento $\Pi K\xi_i - \bigcup_{i=1}^{\ell} D_i$.

Con estas consideraciones el algoritmo toma la siguiente forma:

Algoritmo 4

Paso 0 (de inicialización)

$D^* = \emptyset$

$F = 0$

Paso 1

Calculamos $(i_1, j_1)$ cumpliendo con:

$$U(D^* \cup \{i_1, j_1\}) = \max_{i \in \{1 \ldots p\}} \max_{j \in K\xi_i} U(D^* \cup \{i,j\})$$

Paso 2

Si $U(D^* \cup \{i_1, j_1\}) \geq U(D)$ ir a paso 3.

$U(D^* \cup \{i_1, j_1\}) < U(D)$ ir a paso 4.

Paso 3

$D^* = D^* \cup \{i,j\}$

ir a paso 1

Paso 4

$F = F + 1$

$D^*_F = D^*$

Paso 5

Si $f(\Pi K\xi_i - D^*) \geq g$ $D^* = \emptyset$

$I = I - \hat{I}(D)$

ir a paso 1

$f(\Pi K\xi_i - D^*) < g$ ir a paso 6

<u>Paso 6</u>

$F = F + 1$

$D_F^* = (D_1, \ldots, D_{F-1})^C$

<u>Paso 7</u>

$(D_1, \ldots, D_F)$ es una familia $\mathcal{D}$ terminal para U, parar

83

BIBLIOGRAFIA

Bibliografía

|1| Anderberg, M.R. (1973) "Cluster analysis for applications" Academic Press, New York.

|2| Andersen, E.B. (1980) "Discrete statistical models with social science applications", North-Holland publishing Co.

|3| Ash, R. (1965) "Information Theory", John Wiley.

|4| Bahadur, R.R. (1961) "A representation of the joint distribution of responses to n dichotomous items" de "Studies in item analysis and prediction". Ed. H. Solomon, Standford University Press.

|5| Benzecri, J.P. et Coll (1973) "L'analyse des données", 2 tomos Dunod, 2-eme edit. 1976.

|6| Benzecri, J.P. (1979) "La place de l'apriori", de "sesion d'analyse des donnés" INSEE,IUSP.

|7| Bishop, Y.M.M.; Fienberg, S.E.; Holland, P.W. (1978) "Discrete multivariate analysis. Theory and practice", The MIT Press.

|8| Cochran, W.; Hopkins, C.E. (1961) "Some classification problems with multivariate qualitative data", Biometrics, 17, 10-32.

|9| Diday, E. (1980) "Data analysis and informatics", de "2nd. international Symposium on data analysis and informatics"; ed: North Holland publishing Co.

|10| Diday, E. (1982) "Elements d'analyse des données", Dunod.

|11| Dillon, W.R.; Goldstein, M. (1978) "Discrete discriminant analysis". John Wiley.

|12| Dillon, W.R.; Goldstein, M. (1978) "On the performance of some multinomial classification rules", JASA, vol. 73, n° 362.

|13| Gilbert, E.S., (1968) "On discriminant using qualitative variables". JASA, 63, 1399-1412.

|14| Goodman, L.A. (1978) "Analyzing Qualitative/Categorical Data" Addison-Wesley publishing Co.

|15| Gupta, R.P. (1980) "On a multivariate statistical classification model" de "Multivariate statistical analysis" Ed. R.P. Gupta, North Holland publishing Co.

|16| Haberman, S.J. (1978, 1979) "Analysis of qualitative data" Vol 1: Introductory topics, vol. 2: New developments, Ac. Press.

|17| Hartigan, J.A., (1975) "Clustering algorithms" Wiley.

|18| Hills, M. (1967) "Discrimination and allocation with discrete data" App. Stat. 16, 237-250.

|19| Jambu, M. (1978) "Classification automatique pour l'analyse des données" Tomo 1: méthodes et algorithmes, DUNOD.

|20| Jambu, M.; Lebeaux, M.O., (1978) "Classification automatique pour l'analyse des données" Tomo 2: "Logiciels", DUNOD.

|21| Kass, G.V. (1975) "Significance testing in automatic interaction detection" appl. Stat., 24, 178-189.

|22| Kass, G.V. (1980) "An exploratory technique for investigating large quantities of categorical data" App. Stat., 29, 119-127.

|23| Kulback, S. (1959) "Information theory and statistics" Wiley

|24| Lachenbruch, P.A.; Goldstein, M. (1979) "Discriminant analysis", Biometrics 35, 69-85.

|25| Lachin, J.M. (1973), "On a stepuwise procedure for two popula tions: Bayes decision rules using discrete variables", Biometrics 29, 551-564.

|26| Lebart, L.; Fenelon, J.P. (1971), "Statistique et informati que appliquées". Dunod.

|27| Lebart, L.; Morineau, A.; Tabard, N. (1977) "Techniques de la description statistique", Dunod.

|28| Lebart, L.; Morineau, A.; Fenelon, J.P. (1979) "Traitement des données statistiques", Dunod.

|29| Martin, D.C.; Bradley, R.A. (1972) "Probability models, estimation and classification for multivariate dichotomous populations" Biometrics, 28, 203-222.

|30| Matusita, K. (1955) "On estimation by the minimum distance methods", Annals of the institute of statistical Mathema tics 7, 67-77.

|31| Messenger, R.C.; Mandell, L. (1972) "A model search technique for predictive nominal scale multivariate analysis" JASA, 67, n° 430.

|32| Morgan, J.N.; Messenger, R.C. (1972) "THAID, a sequential ana lysis program for nominal dependent variables", Survey Research centre, I.S.R., University of Michigan.

|33| Ott, J.; Kronmal, R.A. (1976) "Some classification procedures for binary data using orthogonal functions", JASA, 71, 391-399.

|34| Reynolds, H.T. (1977) "The analysis of cross classifications", the free Press.

|35| Romeder, J.M. (1973) "Méthodes et programmes d'analyse discri minante". Dunod.

|36| Sánchez García, M. (1978) "Predicción en variables cualitati vas, aplicación a la intención de voto", Monografía financiada por C.I.S.

|37| Sánchez García, M. (1980) "Tipología y clasificación" Actas de la XII reunión de la SEIO (JACA), por publicar.

|38| Saporta, G. (1977), "Une méthode et un programme d'analyse discriminante pas a pas sur variables qualitatives", de Analyse des données et informàtique, sept. (1977), Versailles.

|39| Saporta, G. (1979) "Une methodologie de discrimination sur variables qualitatives" Congress AFCET-INRIA, 'Reconnai ssance des formes et intelligence artificielle' Toulouse, Sept. 1979.

|40| Saporta, G. (1981) "Un example de discrimination sur variables qualitatives: Les risques presentes par jeunes con ducteurs en assurance automobile", de Multidimensional Data Analysis, Dubrovnik, oct., 1981.

|41| Sonquist, J.N.; Baker, E.C.; Morgan, J.A. (1971) "Searching for structure (alias -AID-III) Michigan institute for social research, University of Michigan.

|42| Van Ryzin J. ed. (1977) "Classification and Clustering", Ac. Press.