

## PROBLEMAS FINITOS DE DECISION CON CONOCIMIENTO CUALITATIVO DE LA DISTRIBUCION A PRIORI

*José A. Cristóbal*  
*Dto. de Estadística*  
*Universidad de Santiago*

### SUMMARY

Decision problems where the set of states of Nature is finite and which has a partial order relationship, are being studied. The distribution "a priori" over the states is unknown, although one has the information which conserves the previous order. Such a distribution is estimated to be subject to a restriction of monotony, and some properties of the estimator are observed.

In a special way this paper deals with the situation where the states are two-dimensional and present their natural order. Finally an algorithm is given to find the quoted estimator in this last case.

### 1. INTRODUCCION

En lo que sigue, abordaremos problemas de decisión en ambiente de riesgo y con experimentación, en donde la distribución de probabilidad a priori sobre los estados de la Naturaleza no es perfectamente conocida, sino que solamente se posee una información cualitativa de la misma. Más concretamente, dados dos estados de la Naturaleza cualesquiera, se conoce a lo más, cuál de ellos es más probable que el otro, si bien no se tiene una idea cuantitativa de esta diferencia de probabilidades.

Se tiene así un planteamiento del problema que podemos enunciar en los siguientes términos:

(\*) Recibido, Mayo, 1981

- Un conjunto de estados de la Naturaleza  $\Omega = \{\theta_1, \theta_2, \dots, \theta_n\}$  finito.
- Un conjunto de acciones puras  $A$
- Una función de pérdida  $L$  definida sobre  $\Omega \times A$
- Un espacio probabilístico  $(\Omega, \mathbf{P}(\Omega), P)$ , donde  $P$  es una distribución a priori sobre  $\Omega$  que se supondrá desconocida, y  $\mathbf{P}(\Omega)$  representa el conjunto de las partes de  $\Omega$ .
- Un preorden parcial  $\lesssim$  sobre  $\Omega$  compatible con  $P$ , entendiendo por tal aquél que verifica:

$$\theta_i \lesssim \theta_j \Rightarrow P(\theta_i) \leq P(\theta_j)$$

En este caso también se dice que la función  $P$  es isótoma respecto al preorden  $\lesssim$ .

- Una variable aleatoria  $\xi$  sobre  $(\Omega, \mathbf{P}(\Omega))$  definida por:  $\xi(\theta_i) = i$ , ( $i = 1, \dots, n$ ), de la cual se ha extraído una muestra aleatoria simple  $x = (x_1, \dots, x_k)$ .

El objetivo consiste en la obtención de aquellas acciones  $a \in A$  que son soluciones de Bayes frente a una distribución  $P^*$  que sea la que más se aproxime, en el sentido de mínimos cuadrados, a la distribución estimada  $\hat{P}$  a partir de  $x$  por el método de máxima verosimilitud, de entre todas las isótonas respecto al preorden  $\lesssim$  dado.

*1.1. PROPOSICION.*— La distribución  $P^*$  más próxima a  $\hat{P}$  en el sentido de mínimos cuadrados entre las isótonas con respecto a  $\lesssim$ , es la función de regresión isotónica de  $\hat{P}$  con pesos unidad.

*Demostración:* En efecto, es ya conocido (Ayer y otros (1955)) que la función de regresión isotónica  $P^*$  minimiza:

$$\sum_{i=1}^n (P_i^* - \hat{P}_i)^2 = \sum_{i=1}^n [P_i^* - (n_i/k)]^2$$

donde

$$P_i^* = P^*(\theta_i), \hat{P}_i = \hat{P}(\theta_i), n_i = \text{Card} \{x_j = i\}_{j=1}^k \quad (i = 1, \dots, n)$$

entre todas las funciones sobre  $\Omega$  que son isótonas respecto al preorden  $\lesssim$  dado.

Por consiguiente sólo falta demostrar que  $P^*$  es a su vez una función de probabilidad sobre  $\Omega$ . Para ello, basta ver que, cuando se utiliza el algoritmo de los conjuntos inferiores para el cálculo de la función de regresión isotónica (ver Barlow y otros (1972)), se realiza una partición del conjunto  $\{1, 2, \dots, n\}$  en bloques solución  $B_j$ , ( $j = 1, \dots, r$ ) dentro de las cuales  $P^*$  es constante y vale:

$$P_i^* = \frac{\sum_{i \in B_j} \hat{P}_i}{\text{Card } B_j}, \quad \forall i \in B_j, (j = 1, \dots, r) \quad (1.1)$$

se tiene así que:

$$\hat{P}_i \geq 0, \quad \forall i \Rightarrow P_i^* \geq 0, \quad \forall i \in \{1, \dots, n\}$$

$$\sum_{i=1}^n P_i^* = \sum_{j=1}^r \sum_{i \in B_j} P_i^* = \sum_{j=1}^r \text{Card } B_j \frac{\sum_{i \in B_j} \hat{P}_i}{\text{Card } B_j} = \sum_{i=1}^n \hat{P}_i = 1 \quad \square$$

1.2. *COROLARIO.*— La distribución  $P^*$  anterior es un estimador de  $P$  mejor que  $\hat{P}$  en el sentido de mínimos cuadrados.

*Demostración:* Representaremos por  $P = (P_1, \dots, P_n)$  la verdadera distribución de probabilidad a priori sobre  $\Omega$ . La función de regresión isotónica de  $\hat{P}$  respecto a  $\lesssim$  verifica (ver Van Eeden (1956)):

$$\sum_{i=1}^n (\hat{P}_i - f_i)^2 \geq \sum_{i=1}^n (\hat{P}_i - P_i^*)^2 + \sum_{i=1}^n (P_i^* - f_i)^2 \quad (1.3)$$

para cualquier  $f = (f_1, \dots, f_n)$  que sea isótoma con respecto a  $\lesssim$ . En particular, sabemos que  $P$  es isótoma respecto a  $\lesssim$ , luego se tiene:

$$\begin{aligned} \sum_{i=1}^n (\hat{P}_i - P_i)^2 &\geq \sum_{i=1}^n (\hat{P}_i - P_i^*)^2 + \sum_{i=1}^n (P_i^* - P_i)^2 \geq \\ &\geq \sum_{i=1}^n (P_i^* - P_i)^2 \quad \square \end{aligned}$$

Este corolario señala la importancia de escoger  $P^*$  en lugar de  $\hat{P}$  para estimar  $P$  cuando se conoce un preorden con el cual esta última es compatible.

1.3. *Aplicación al caso en que la distribución a priori sigue un esquema de contagio de Polya de parámetros desconocidos.* — Suponemos un esquema de contagio de Polya cuyos elementos iniciales son  $B$  y  $N$ , y donde representaremos con  $c$  el número de elementos que se añaden al sistema en cada repetición independiente. Si llamamos  $\theta_i$ , ( $i = 0, 1, \dots, n$ ) al resultado (elementos del primer tipo obtenidos en las  $n$  repeticiones), se obtiene:

$$P(\theta_0) = \frac{\prod_{\ell=0}^{n-1} (N + \ell c)}{\prod_{\ell=0}^{n-1} (B + N + \ell c)}, \quad P(\theta_n) = \frac{\prod_{\ell=0}^{n-1} (B + \ell c)}{\prod_{\ell=0}^{n-1} (B + N + \ell c)}$$

$$P(\theta_k) = \frac{\prod_{\ell=0}^{n-k-1} (N + \ell c) \prod_{\ell=0}^{k-1} (B + \ell c)}{\prod_{\ell=0}^{n-1} (B + N + \ell c)}, \quad (1 \leq k \leq n-1) \quad (1.4)$$

de donde, mediante un sencillo cálculo se obtiene que los elementos  $P(\theta_i)$ , ( $i = 0, 1, \dots, n$ ) son no decrecientes con el subíndice  $i$  si y solo si se verifican todas las desigualdades siguientes:

$$B - N \geq (n - 2k - 1)c, \quad (k = 0, 1, \dots, n-1) \quad (1.5)$$

para lo cual, basta con que se cumpla:

$$B - N \geq (n - 1)c \quad (1.6)$$

De este modo, si  $B$  y  $N$  son desconocidos, pero su diferencia está acotada de forma que se verifica (1.6), se tiene que la distribución  $P$  anterior sobre  $\Omega = \{\theta_0, \dots, \theta_n\}$  no es conocida cuantitativamente, pero es isótoma con respecto al orden (total):

$$\theta_0 \lesssim \theta_1 \lesssim \dots \lesssim \theta_n$$

A continuación, supongamos que se realiza el experimento descrito un cierto número de veces de manera independiente, obteniéndose una muestra aleatoria simple que permite calcular el estimador  $\hat{P}$  de la distribución  $P$ . Al hallar la regresión isotónica  $P^*$ , el conjunto  $\{0, 1, \dots, n\}$  queda dividido en una colección de bloques  $B_j$ , ( $j = 1, \dots, r$ ) dentro de los cuales  $P^*$  es constante.

Sea por fin el problema de decisión consistente en pronosticar el resultado  $\theta_i \in \Omega$  de una nueva repetición del experimento. Se tiene así un conjunto de acciones  $A = \{a_i\}_{i=0}^n$  donde  $a_i$  representa pronosticar el resultado  $\theta_i$ , ( $i = 0, \dots, n$ ). Si la función de pérdida viene expresada por el error cuadrático:

$$L(a_s, \theta_i) = (s - i)^2 \quad , \quad (s, i = 0, \dots, n) \quad (1.7)$$

entonces el riesgo de la acción  $a_s$  frente a la distribución  $P^*$  sobre  $\Omega$  viene dado por:

$$R^*(a_s) = \sum_{i=0}^n (s - i)^2 P_i^* = \sum_{j=1}^r \sum_{i \in B_j} (s - i)^2 P_i^*$$

por lo que, si el bloque  $B_j = \{\alpha_j, \dots, \alpha_j + \beta_j\}$ , ( $j = 1, \dots, r$ ), se tiene:

$$R^*(a_s) = \sum_{j=1}^r \frac{\hat{P}(B_j)}{\beta_j + 1} \sum_{i=\alpha_j}^{\alpha_j + \beta_j} (s - i)^2 \quad (1.8)$$

Como el objetivo es minimizar este riesgo, podemos considerar  $s$  como una variable continua en (1.8), e igualando a cero la correspondiente derivada resulta:

$$s = \sum_{j=1}^r \frac{2\alpha_j + \beta_j}{2} \hat{P}(B_j) \quad (1.9)$$

con lo cual la acción óptima corresponde a un valor de  $s$  que es el entero anterior o posterior al valor dado en (1.9). Obsérvese que el coeficiente de cada  $\hat{P}(B_j)$  es la media aritmética de los extremos de tal bloque  $B_j$ , con lo cual, (1.9) puede reescribirse:

$$s = \sum_{j=1}^r \frac{\sum_{\ell=0}^{\beta_j} (\alpha_j + \ell)}{\beta_j + 1} \hat{P}(B_j) = \sum_{j=1}^r \sum_{\ell=0}^{\beta_j} (\alpha_j + \ell) P_j^*$$

es decir:

$$s = \sum_{i=0}^n i P_i^* \quad (1.10)$$

y coincide por lo tanto con la media de la distribución  $P^*$ , como era de esperar ya que la función de pérdida tomada ha sido el error cuadrático.

Si el valor de  $s$  no es entero, podemos aleatorizar el espacio de acciones  $A$  y considerar la correspondiente perspectiva aleatoria entre  $a_{[s]}$  y  $a_{[s]+1}$  (donde  $[s]$  expresa la parte entera de  $s$ ) de modo que el riesgo de Bayes sea alcanzable, en cuyo caso sabemos que coincide con la varianza de la distribución  $P^*$ .

## 2. Un caso notable de orden parcial sobre $\Omega$

En muchos casos prácticos, como pueden ser problemas de estimación paramétrica ordenados, los elementos de  $\Omega$  se encuentran relacionados por un orden que es completo. En tales situaciones el modo de operar es inmediato, pues una vez obtenida  $\hat{P}$  basta aplicar un algoritmo, como puede ser el de fusión de elementos adyacentes (ver Ayer y otros (1955)), para calcular  $P^*$ , y a partir de ahí el problema se convierte en uno usual de decisión.

Oras muchas veces, aunque el preorden sobre  $\Omega$  no sea completo, es de una forma especialmente importante, a saber: los elementos de  $\Omega$  son bidimensionales, esto es,  $\Omega$  se puede expresar como un cierto producto cartesiano  $\Omega = \Omega_1 \times \Omega_2$ , donde  $\Omega_1$  y  $\Omega_2$  son conjuntos finitos, y dentro del cual los elementos se encuentran relacionados por el orden natural:

$$(\theta_1, \theta_2) \succeq (\theta'_1, \theta'_2) \Leftrightarrow \theta_1 \geq \theta'_1 \text{ y } \theta_2 \geq \theta'_2 \quad (2.1)$$

Este orden es sólo parcial, pero da a  $\Omega$  una estructura de retículo distributivo cuyas propiedades utilizamos más tarde.

*2.1. EJEMPLO.*— Supongamos que un individuo tiene una cierta característica  $c_1$  y engendra hijos que tienen dicha característica con probabilidad  $\alpha_1$ , los cuales a su vez se encuentran en la misma situación; sin embargo si un hijo cualquiera no presenta  $c_1$  ninguno de sus descendientes la presenta. A la vez, se plantea idéntica situación para una característica  $c_2$  que se transmite con probabilidad  $\alpha_2$ . Los valores de  $\alpha_1$  y  $\alpha_2$

no se conocen, pero se puede asegurar que están acotados superiormente por  $\sqrt{2}/2$ .

Denominamos  $\theta_1^{(1)}$  y  $\theta_2^{(1)}$  a los sucesos de que un miembro de la tercera generación respecto del primer individuo, elegido al azar, posea  $c_1$  o no, respectivamente, y por  $\theta_1^{(2)}$  y  $\theta_2^{(2)}$  los análogos respecto a  $c_2$ . Se tiene entonces:

$$P_1(\theta_1^{(1)}) \leq P_1(\theta_2^{(1)}) \Leftrightarrow \alpha_1^2 \leq \alpha_1(1 - \alpha_1) + (1 - \alpha_1) \Leftrightarrow \alpha_1 \leq \sqrt{2}/2$$

$$P_2(\theta_1^{(2)}) \leq P_2(\theta_2^{(2)}) \Leftrightarrow \alpha_2^2 \leq \alpha_2(1 - \alpha_2) + (1 - \alpha_2) \Leftrightarrow \alpha_2 \leq \sqrt{2}/2$$

las cuales se verifican, debido a la hipótesis de acotación realizada.

Si tenemos en cuenta a la vez las dos características  $c_1$  y  $c_2$ , se tienen cuatro estados posibles:

$$\theta_{ij} = (\theta_i^{(1)}, \theta_j^{(2)}) \quad (i, j = 1, 2)$$

los cuales se encuentran relacionados por el orden natural dado en (2.1), que, por construcción, es compatible con la función de probabilidad producto  $P = P_1 \times P_2$  sobre  $\Omega = \{\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}\}$ .

Como en esta nueva situación el orden es parcial los algoritmos existentes para el cálculo de la función de regresión isotónica son mucho más complicados, si bien en el párrafo 3 describimos uno para este tipo de orden particular.

En lo que sigue, estudiamos una sencilla representación gráfica de un problema de decisión, cuando  $\Omega = \Omega_1 \times \Omega_2$ , en donde  $\Omega_1$  y  $\Omega_2$  tienen sólo dos elementos cada uno (como en el ejemplo 2.1). La generalización al caso en que  $\Omega_1$  tenga dos elementos y  $\Omega_2$  tenga tres, o bien que ambos posean tres elementos, es inmediata; en dimensión superior no se puede visualizar.

**2.2. DEFINICION.**— Dados los puntos  $Q_i = (x_i, y_i)$ , ( $i = 1, 2$ ) en  $\mathbb{R}^2$  y los haces de rectas:  $r_i: a_i x + b_i y = \text{cte.}$  ( $i = 1, 2$ ), se llama *mezcla* de  $Q_1$  y  $Q_2$  con respecto a  $r_1$  y  $r_2$ , al punto

$$Q_m = \left( \frac{a_1 x_1 + b_1 y_1}{a_1 + b_1}, \frac{a_2 x_2 + b_2 y_2}{a_2 + b_2} \right) \quad (2.2)$$

Gráficamente (ver fig. 2.1) este punto se obtiene de la manera siguiente: se traza la recta de  $r_1$  que pasa por  $Q_1$  y se corta con la bisectriz del primer y tercer cuadrante, trazando la vertical  $\nu$  que pasa por este punto. Luego se realiza operación análoga con  $r_2$  pero esta vez trazando la horizontal  $h$  por el punto de corte con la bisectriz. El punto  $Q_m$  es precisamente la intersección de  $\nu$  con  $h$ , como fácilmente se puede comprobar.

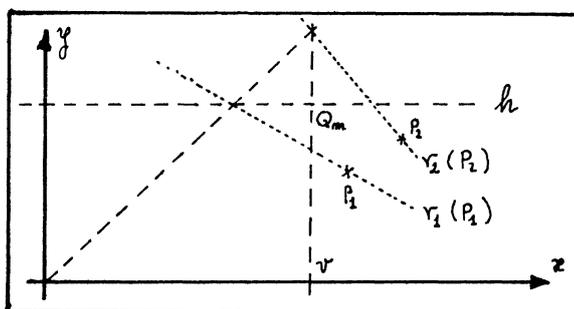


Fig. 2.1

2.3. DEFINICION.— Dados  $A_1, A_2$ , subconjuntos de  $\mathbb{R}^2$ , una correspondencia  $f: A_1 \rightarrow A_2$  entre ambas, (con  $A_2 = f(A_1)$ ), y los haces de rectas  $r_i: a_i x + b_i y = cte$ , ( $i = 1, 2$ ), se llama *mezcla* de  $A_1$  y  $A_2$  respecto a  $f, r_1$  y  $r_2$ , al subconjunto de  $\mathbb{R}^2$  formado por las mezclas de  $Q_1$  y cada  $Q_2 \in f(Q_1)$  respecto a  $r_1$  y  $r_2$ , cuando  $Q_1$  recorre todo  $A_1$ .

Sea entonces un problema de decisión con espacio de estados  $\Omega = \{\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}\}$ , espacio de acciones  $A = \{a_i\}_{i \in I}$  y función de pérdida  $L(a_i, \theta_{jk}) = L_{jk}^{(i)}$ , ( $i \in I; j, k = 1, 2$ ).

Realizamos la siguiente representación gráfica: a cada acción  $a_i$  le hacemos corresponder el par de puntos en el plano  $Q_1^{(i)} = (L_{11}^{(i)}, L_{12}^{(i)})$  y  $Q_2^{(i)} = (L_{21}^{(i)}, L_{22}^{(i)})$ . Además, si llamamos  $P_{ij} = P(\theta_{ij})$  para una cierta distribución de probabilidades  $P$  sobre  $\Omega$ , y formamos los haces de rectas:

$$r_i: \frac{P_{i1}}{P_i} x + \frac{P_{i2}}{P_i} y = cte, \quad (i = 1, 2) \quad (2.3)$$

en donde  $P_i = \sum_{j=1}^2 P_{ij}$ , ( $i = 1, 2$ ) son las probabilidades marginales fila, a cada acción  $a_i$  se le puede hacer corresponder el punto del plano  $Q_m^{(i)}$  mezcla de  $Q_1^{(i)}$  y  $Q_2^{(i)}$  respecto de los haces de rectas dados en (2.3).

Se verifica entonces el siguiente resultado:

**2.4. TEOREMA.**— Dada una distribución de probabilidad  $P = (P_{11}, P_{12}, P_{21}, P_{22})$  sobre  $\Omega$ , la acción (o acciones) de Bayes viene caracterizada en una representación gráfica de las pérdidas por el punto en que primero toca el haz de rectas:  $p_1x + p_2y = \text{cte}$  (al aumentar la cte) al conjunto mezcla de  $\{(L_{11}^{(i)}, L_{12}^{(i)})\}_{i \in I}$  y  $\{(L_{21}^{(i)}, L_{22}^{(i)})\}_{i \in I}$  respecto de la correspondencia que conserva el superíndice y de los haces de rectas de definidos en (2.3).

*Demostración:* El riesgo de la acción  $a_i$ , ( $i \in I$ ) viene dado por:

$$R_i = \sum_{j,k} P_{jk} L_{jk}^{(i)} = \sum_{j=1}^2 P_j \left( \frac{P_{j1}}{P_j} L_{j1}^{(i)} + \frac{P_{j2}}{P_j} L_{j2}^{(i)} \right) \quad (2.4)$$

y por lo tanto el problema es equivalente a otro en el que  $\Omega' = \{\theta'_1, \theta'_2\}$  con distribución de probabilidad  $(P_1, P_2)$ , y con pérdidas:

$$L(a_i, \theta'_j) = \frac{P_{j1}}{P_j} L_{j1}^{(i)} + \frac{P_{j2}}{P_j} L_{j2}^{(i)}, \quad (j = 1, 2) \quad (2.5)$$

Ahora bien, de las definiciones 2.2 y 2.3 se desprende que este conjunto de pérdidas viene representado en  $\mathbb{R}^2$  por la mezcla enunciada en el teorema. El punto solución de Bayes de este nuevo problema será la mezcla de dos puntos  $(L_{11}^{(i)}, L_{12}^{(i)})$  y  $(L_{21}^{(i)}, L_{22}^{(i)})$  que representan a  $a_i$ , la solución Bayes del problema original.  $\square$

Este teorema 2.4 nos da un sugestivo método gráfico para la obtención de la solución de Bayes en problemas de decisión cuyo espacio de estados es del tipo especial que consideramos, en base a la utilización de distribuciones marginales y condicionadas.

Observemos también que la validez del mismo se extiende a cualquier tipo de distribuciones de probabilidad sobre  $\Omega$ , si bien en el caso que nos ocupa estamos especialmente interesados en distribuciones que son isótonas respecto al orden parcial (2.1) definido sobre  $\Omega$ .

2.5. *PROPOSICION.*— Si una acción  $a_i$  es la mixtura de otras dos con pesos  $\alpha$  y  $1 - \alpha$  respectivamente ( $0 < \alpha < 1$ ), lo mismo le sucede a los correspondientes puntos de  $\mathbb{R}^2$  en la representación gráfica de las pérdidas.

*Demostración:* Sean  $a_1, a_2 \in A$ , y llamemos  $a_\alpha = \alpha a_1 + (1 - \alpha) a_2$ . Entonces su pérdida vale:

$$L(a_\alpha, \theta_{ij}) = \alpha L(a_1, \theta_{ij}) + (1 - \alpha) L(a_2, \theta_{ij}), \quad (i, j = 1, 2) \quad (2.6)$$

y en cuanto a su riesgo, toma el siguiente valor:

$$R_\alpha = \sum_{i,j} P_{ij} L(a_\alpha, \theta_{ij}) = \sum_{i=1}^2 P_i \sum_{j=1}^2 \frac{P_{ij}}{P_i} L(a_\alpha, \theta_{ij}) \quad (2.7)$$

Llevando ahora (2.6) a la expresión (2.7) se obtiene:

$$R_\alpha = \sum_{i=1}^2 P_i \left[ \alpha \sum_{j=1}^2 \frac{P_{ij}}{P_i} L(a_1, \theta_{ij}) + (1 - \alpha) \sum_{j=1}^2 \frac{P_{ij}}{P_i} L(a_2, \theta_{ij}) \right] \quad (2.8) \quad \square$$

2.6. *COROLARIO.*— Si el espacio de acciones  $A$  es finito, el subconjunto de  $\mathbb{R}^2$  correspondiente a  $A^*$  (espacio de acciones aleatorizadas) en la representación gráfica de las pérdidas es un polítopo convexo (clausura convexa de los puntos correspondientes a cada acción pura).

*Demostración:* Es inmediata a partir de la proposición 2.5 y de su inmediata generalización al caso de más de dos acciones.  $\square$

2.7. *EJEMPLO.*— Sea  $\Omega = \{\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}\}$  con  $A = \{a_1, a_2, a_3\}$  y función de pérdida:  $L(a_1) = (2, 4, 5, 2)$ ,  $L(a_2) = (1, 6, 3, 3)$ ,  $L(a_3) = (3, 5, 6, 3)$  siendo la distribución sobre  $\Omega$  la  $P = (0.1, 0.3, 0.2, 0.4)$ . Entonces la representación gráfica de las pérdidas para cada una de las tres acciones puras se puede observar en la figura 2.2(a).

Los puntos mezcla que corresponden a cada una de ellas valen respectivamente:

$$Q_m^{(1)} = (3.5, 3) \quad Q_m^{(2)} = (4.75, 3) \quad Q_m^{(3)} = (4.5, 4)$$

El haz de rectas que hay que trasladar es:

$$0,4x + 0,6y = \text{cte}$$

que toca por primera vez al conjunto mezcla en el punto  $Q_m^{(1)}$ , alcanzando entonces la constante un valor de 3.2. (Ver figura 2.2(b)).

Se tiene así que la acción de Bayes es  $a_1$  y su riesgo vale  $R^*(a_1) = 3,2$ .

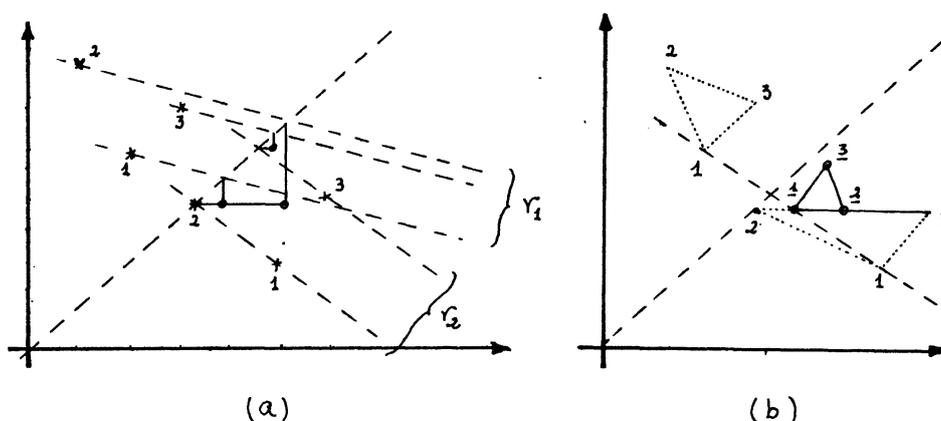


Fig. 2.2

### 3. Algoritmo para la regresión isotónica sobre el orden natural en el plano

Ya hemos visto en el párrafo 2 la importancia considerable que tiene considerar el espacio base (conjunto de estados de la Naturaleza en los problemas de decisión planteados) como un producto cartesiano  $\Omega = \Omega_1 \times \Omega_2$ , donde  $\Omega_1$  y  $\Omega_2$  son conjuntos finitos y dentro del cual consideramos el orden usual de  $\mathbb{R}^2$  que en este caso es solo parcial.

Aunque en todo lo que sigue sólo consideramos el caso bidimensional, la extensión al  $n$ -dimensional para  $n \geq 3$  es inmediata, si bien la nomenclatura se amplía algo, y el número de operaciones a realizar se aumenta de modo considerable al crecer el valor de  $n$ .

Desde luego, existen algunos algoritmos para el cálculo de la función de regresión isotónica cuando el espacio base tiene un orden parcial cualquiera, pero su aplicación práctica queda reducida a casos en que  $\Omega$  es extremadamente pequeño o bien el orden sobre él es excesivamente simple, (ver Brunk (1955), Brunk, Ewing y Utz (1957), Miles (1959), Van Eeden (1957 y 1958)). Por otra parte, Gebhardt (1970) construyó un algoritmo para el caso planteado en el presente párrafo el cual tiene ya una aplicación clara en conjuntos base relativamente grandes, si bien su principal inconveniente estriba en que en uno de los pasos de su algoritmo hay que observar si cada bloque en estudio es “descomponible” o no (según la nomenclatura de su autor), para lo cual es necesario otro algoritmo auxiliar, que también incluye Gebhardt en el mismo artículo. En suma, el algoritmo total planteado por él está basado en un gran número de comparaciones y trasvase de elementos para reordenarlos constantemente, lo que hace elevarse en gran modo el tiempo de cómputo del ordenador con que se trabaje. Por contra, el algoritmo que presentamos en el presente trabajo está basado en operaciones básicas (sumas y divisiones), realizando un número muy pequeño de comparaciones y ningún reordenamiento, lo que le hace más práctico, y ante todo más rápido.

En lo que sigue, denominaremos “bloque inferior” a cualquier subconjunto de  $\Omega$  que verifique que si un elemento está en el bloque, todos los que le preceden también lo están.

3.1. *TEOREMA.*— Todo bloque inferior en  $(\Omega = \prod_{i=1}^2 \Omega_i, \lesssim)$  con  $\Omega_1 = \{1 \dots R\}$ ,  $\Omega_2 = \{1 \dots T\}$  viene dado por una unión

$$\bigcup_{\alpha=1}^k C(I_\alpha, J_\alpha)$$

donde  $I_\alpha \in \Omega_1$ ,  $J_\alpha \in \Omega_2$ ,  $C(I_\alpha, J_\alpha) = \{1, \dots, I_\alpha\} \times \{1, \dots, J_\alpha\}$  es el conjunto dominado por el par  $(I_\alpha, J_\alpha)$ , y  $k$  es un entero cualquiera comprendido entre 1 y el mínimo de  $R$  y  $T$ .

*Demostración:* De la definición de bloque inferior se deduce que para cualquier elemento suyo  $(I_\alpha, J_\alpha)$  debe ocurrir que todos los que le preceden están en el bloque, pero precisamente tal conjunto es  $C(I_\alpha, J_\alpha)$ . De la finitud de  $\Omega$  se deduce ya el enunciado.  $\square$

Obsérvese que estos conjuntos se pueden expresar más explícitamente en la forma (ver figura 3.1)

$$\bigcup_{\alpha=1}^k C(I_\alpha, J_\alpha) = \left\{ \begin{array}{l} I = 1 \dots I_1 \\ J = 1 \dots J_1 \end{array} \right\} \cup \left\{ \begin{array}{l} I = I_1 + 1, \dots, I_2 \\ J = 1, \dots, J_2 \end{array} \right\} \cup \dots \cup \left\{ \begin{array}{l} I = I_{k-1} + 1, \dots, I_k \\ J = 1, \dots, J_k \end{array} \right\} \quad (3.1)$$

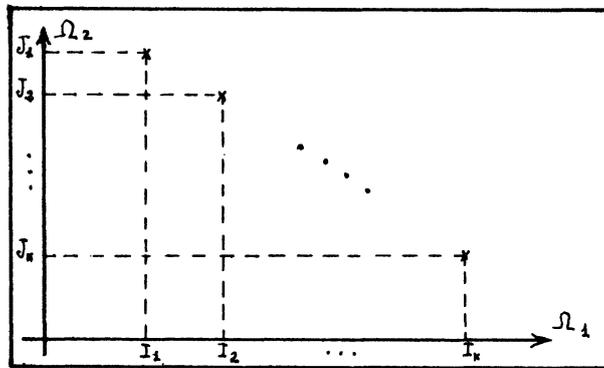


Fig. 3.1

en donde las uniones del miembro derecho son disjuntas. Notemos también que se ha descartado el caso trivial en que haya un  $(I_\alpha, J_\alpha)$  dentro de un  $C(I_{\alpha'}, J_{\alpha'})$ , ya que entonces la unión

$$C(I_\alpha, J_\alpha) \cup C(I_{\alpha'}, J_{\alpha'}) = C(I_{\alpha'}, J_{\alpha'})$$

y por consiguiente el elemento  $(I_\alpha, J_\alpha)$  se puede eliminar de la expresión primera.

A continuación trataremos con funciones  $G(I, J)$  definidas sobre  $\Omega = \Omega_1 \times \Omega_2$  o sobre una parte estrictamente contenida en él. Como en gran número de casos se da esta última situación, definimos a partir de  $G$  un par de funciones:

$$G'(I, J) = \begin{cases} G(I, J) & \text{si } (I, J) \text{ pertenece al dominio de } G \\ 0 & \text{en caso contrario} \end{cases} \quad (3.2)$$

$$W(I, J) = \begin{cases} 1 & \text{si } (I, J) \text{ pertenece al dominio de } G \\ 0 & \text{en caso contrario} \end{cases}$$

Observemos que  $W(I, J)$ , llamada función de pesos, es la función indicadora del dominio de  $G(I, J)$ .

3.2. *TEOREMA.* — Dada la función  $G(I, J)$  definida sobre un subconjunto de  $\Omega$ , la media de  $G$  en el bloque inferior  $\bigcup_{\alpha=1}^k C(I_\alpha, J_\alpha)$  donde los  $(I_\alpha, J_\alpha)$  están ordenados como en la fig. 3.1 viene dada por:

$$P^1(I_1, J_1) = \frac{S(I_1, J_1)}{V(I_1, J_1)} \quad (3.3)$$

$$P^k(I_\alpha, J_\alpha; \alpha = 1, \dots, k) = \frac{\sum_{\alpha=1}^k S(I_\alpha, J_\alpha) - \sum_{\alpha=1}^{k-1} S(I_\alpha, J_{\alpha+1})}{\sum_{\alpha=1}^k V(I_\alpha, J_\alpha) - \sum_{\alpha=1}^{k-1} V(I_\alpha, J_{\alpha+1})} \quad (\text{si } k \geq 2)$$

donde  $S(I_\alpha, J_\alpha) = \sum G'(I, J)$ ,  $V(I_\alpha, J_\alpha) = \sum W(I, J)$ , y donde ambos sumatorios se extienden al conjunto

$$C(I_\alpha, J_\alpha) = \left\{ \begin{array}{l} I = 1, \dots, I_\alpha \\ J = 1, \dots, J_\alpha \end{array} \right\}$$

En el caso indeterminado en que tanto el numerador como el denominador se anulen, el valor de  $P$  es cero (el denominador sólo no puede anularse).

*Demostración:* Para  $k = 1$ , el valor de  $P^1$  es inmediato de su definición.

Para  $k = 2$ , el valor de  $P^2$  es el cociente entre la suma de todos los  $G(I, J)$  del bloque dividido por el número de elementos del mismo en que está definida.

Se tiene así que el numerador es:

$$\sum_{I, J=1}^{I_1, J_1} G'(I, J) - \sum_{I, J=1}^{I_2, J_2} G'(I, J) - \sum_{I, J=1}^{I_1, J_2} G'(I, J)$$

ya que

$$C(I_1, J_1) \cap C(I_2, J_2) = C(I_1, J_2) \quad (3.4)$$

Por tanto dicho numerador queda

$$S(I_1, J_1) + S(I_2, J_2) - S(I_1, J_2)$$

En cuanto al denominador basta razonar igualmente con las funciones de pesos  $W$ .

Por lo demás, la prueba se puede completar por inducción sobre  $k$ , ya que si es cierta la expresión de  $P^k$  para un cierto valor de  $k$ , la media  $P^{k+1}$  para el valor  $k + 1$  tiene por numerador:

$$\begin{aligned} & \left( \sum_{\alpha=1}^k S(I_\alpha, J_\alpha) - \sum_{\alpha=1}^{k-1} S(I_\alpha, J_{\alpha+1}) \right) + S(I_{k+1}, J_{k+1}) - S(I_k, J_{k+1}) = \\ & = \sum_{\alpha=1}^{k+1} S(I_\alpha, J_\alpha) - \sum_{\alpha=1}^k S(I_\alpha, J_{\alpha+1}) \end{aligned}$$

sucediendo la relación análoga para el denominador.  $\square$

### Descripción del algoritmo

El algoritmo queda descompuesto en los siguientes pasos:

- i) Dada  $G(I, J)$  calcular  $G'(I, J)$  y  $W(I, J)$  según 3.2.
- ii) Calcular  $P^1(I, J)$ ,  $\forall (I, J) \in \Omega$  según 3.3.
- iii) Hallar el mínimo de los  $P^1(I, J)$  dentro del conjunto de los pares  $(I, J)$  tales que  $W(I, J) = 1$ . Le llamamos  $M^1(I^1, J^1)$ .
- iv) Para cada  $k$  (con  $k \geq 2$ ), calcular  $P^k(I_\alpha, J_\alpha; \alpha = 1, \dots, k)$  según 3.3 para todos los pares  $(I_\alpha, J_\alpha)$  ( $\alpha = 1, \dots, k$ ) con  $W(I_\alpha, J_\alpha) = 1$  y con la condición:  $I_1 < I_2 < \dots < I_k$  y  $J_1 > J_2 > \dots > J_k$ .
- v) Hallar el mínimo de los  $P^k(I_\alpha, J_\alpha; \alpha = 1, \dots, k)$  dentro del conjunto dado en (iv). Le llamamos  $M^k(I_\alpha^k, J_\alpha^k; \alpha = 1, \dots, k)$ .

- vi) Si  $k < \min(R, T)$  aumentar  $k$  en una unidad y pasar a (iv). En caso contrario hallar el  $\min M^k(I_\alpha^k, J_\alpha^k; \alpha = 1, \dots, k)$ . Le llamamos:

$$M^{k(L)}(I_\alpha^{k(L)}, J_\alpha^{k(L)}; \alpha = 1, \dots, k(L))$$

En la primera iteración,  $L = 1$ .

- vii) Se toma  $G'(I, J) = W(I, J) = 0$  para todo par  $(I, J) \in \bigcup_{\alpha=1}^{k(L)} C(I_\alpha^{k(L)}, J_\alpha^{k(L)})$ .
- viii) Calcular  $W(R, T)$ . Si es nulo el algoritmo se ha acabado. En caso contrario pasar a (ii) aumentando  $L$  en una unidad.

Una vez terminado el algoritmo la función de regresión isotónica toma los valores:

$$G^*(I, J) = M^{k(L)}(I_\alpha^{k(L)}, J_\alpha^{k(L)}; \alpha = 1, \dots, k(L)),$$

$$\forall (I, J) \in \bigcup_{\alpha=1}^{k(L)} C(I_\alpha^{k(L)}, J_\alpha^{k(L)})$$

habiendo tantos bloques solución como iteraciones sobre  $L$  se han realizado.

## REFERENCIAS

- AYER, M., H.D. BRUNK, G.M. EWING, W.T. REID y E. SILVERMAN (1955). "An empirical distribution function for sampling with incomplete information". Ann. Math. Statist. 26, 641-647.
- BARLOW, R.E., D.J. BARTHOLOMEU, J.M. BREMNER y H.D. BRUNK (1972). "Statistical inference under order restrictions". John Wiley & Sons.
- BRUNK, H.D. (1955), "Maximun likelihood estimates of monotone parameters". Ann. Math. Statist. 26, 607-616.
- BRUNK, H.D., G.M. EWING y W.R. UTZ (1957). "Minimizing integrals in certain classes of monotone functions". Pacif. J. Math. 7, 833-847.
- DE GROOT, M.H. (1970). "Optimal statistical decisions". McGraw-Hill.
- GEBHARDT, F. (1970). "An algorithm for monotone regression with one or more independent variables". Biometrika. 57, 263-271.

- MILES, R.E. (1959). "The complete amalgamation into bloks, by weighted means, of a finite set of real numbers". *Biometrika*. 46, 317-327.
- SACKROWITZ, H.B. (1970). "Estimation for ordered parameter sequences: The discrete case". *Ann. Math. Statist.* 41, 609-620.
- VAN EEDEN, C. (1956). "Maximun likelihood estimation of ordered probabilities". *Indag. Math.* 18, 444-455.
- VAN EEDEN, C. (1957). "Maximun likelihood estimation of partially or completely ordered parameters". *Indag. Math.* 19, 128-136 y 201-211.
- VAN EEDEN, C. (1958). "Testing and estimating ordered parameters of probability distributions". Studentendrukkerij Poortpers. Amsterdam.