UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS
Departamento de Estadística e Investigación Operativa I

DISTRIBUCIONES DE MAXIMA ENTROPÍA EN ESPACIOS DE PROBABILIDAD TRANSFORMADOS

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Juan Francisco Serra Cuñat

Bajo la dirección del doctor:
Agustín Turrero Nogués

Madrid, 2006

- **ISBN: 978-84-669-2941-7**

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS
Departamento de Estadística e Investigación Operativa I

TESIS DOCTORAL

Distribuciones de máxima entropía en espacios de probabilidad transformados

Autor: Juan Francisco Serra Cuñat
Director: Agustín Turrero Nogués
Madrid, Febrero de 2006

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS
Departamento de Estadística e Investigación Operativa I

TESIS DOCTORAL

Distribuciones de máxima entropía en espacios de probabilidad transformados

Memoria presentada por D. Juan Francisco Serra Cuñat para optar al grado de Doctor en Ciencias Matemáticas por la Universidad Complutense de Madrid en el programa de tercer ciclo de Estadística e Investigación Operativa.

Realizada bajo la dirección del Dr. D. Agustín Turrero Nogués, profesor Titular del Departamento de Estadística e Investigación Operativa I de la Universidad Complutenese de Madrid.

Autor: Juan Francisco Serra Cuñat
Director: Agustín Turrero Nogués
Madrid, Febrero de 2006

A Francisco, Teresa y María.

A Rafael y Estrella.

Índice general

CONTENIDO Y ESTRUCTURA

El contenido de esta memoria se encuentra estructurado en cinco capítulos. Fundamentalmente todo el esfuerzo se centra en aplicar el principio de máxima entropía (por el cual se elige como distribución teórica, aquella que maximiza la entropía) tras efectuar una transformación lineal de un espacio original de distribuciones de probabilidad discretas finitas. Posteriormente, se utilizarán los resultados obtenidos en un modelo de supervivencia censurado aleatoriamente por la derecha, cuyo espacio de probabilidades se puede obtener, precisamente, mediante una transformación lineal (determinada) del espacio de probabilidades asociado al experimento no censurado.

Por ser las medidas de entropía piezas fundamentales de este trabajo, el capítulo I se dedica a presentar las medidas de información (incertidumbre) denominadas entropías, dándose una visión histórica de su origen y del contexto en el que aparecen, así como de su interpretación en Estadística. Se habla de la entropía de Shannon y de las propiedades que verifica, pasando posteriormente a definir y revisar las medidas generalizadas de entropía propuestas en la literatura.

La Programación Matemática tiene también un papel esencial en este trabajo, el capítulo II está dedicado en su totalidad a revisar los conceptos y técnicas fundamentales que se utilizan en Programación Matemática, en especial en programas no lineales.

El capítulo III comienza con la presentación detallada de la transformación lineal del espacio de probabilidades, a continuación se formula el programa matemático a resolver (de acuerdo con el principio de máxima entropía) y se explica el método a seguir para conseguir una solución aproximada del mismo, solución que posee dos importantes cualidades:

1. Se puede considerar bajo determinadas condiciones como solución del programa matemático citado independientemente de la medida de entropía considerada.

2. Puede servir como punto inicial en los métodos de optimización denominados "métodos de busqueda directa".

Se analizan posteriormente las características de dicha solución así como también se es-

tudia el error cometido en los casos prefijados. Por último se desarrollan detalladamente los cálculos necesarios para conseguir la solución aproximada a partir de las condiciones necesarias y suficientes de Kuhn-Tucker para programas convexos. Cabe destacar que esta solución se obtiene resolviendo sistemas de ecuaciones lineales, lo que facilita considerablemente su cálculo.

En el capítulo IV se introduce el Análisis de Supervivencia. Se explican las características fundamentales de esta parte de la Estadística, se presentan los conceptos de censura, de función de riesgo, etc. y se analizan algunos modelos paramétricos y no paramétricos utilizados para describir el comportamiento de la variable aleatoria no negativa denominada "tiempo de vida".

En el capítulo V se aplican los resultados obtenidos en el capítulo III a un modelo de supervivencia con datos agrupados censurado aleatoriamente por la derecha, y se analizan las características particulares del mismo que hacen que sea un caso particular entre los estudiados en el capítulo III. El capítulo acaba con una aplicación práctica en la que se recogen los resultados para varias distribuciones de censura.

Por último, los apéndices A y B contienen detalladamente todo el soporte algebraico utilizado a lo largo de los capítulos, especialmente del III. El primero de estos apéndices está dedicado a matrices y formas cuadráticas pues constituyen instrumentos imprescindibles en todo el proceso matemático seguido. El segundo contiene una breve introducción a los espacios métricos y normados.

Agradecimientos

Eterna gratitud al Dr. D. Agustín Turrero Nogués por todo el tiempo que me ha dedicado durante la dirección de esta tesis. Agustín ha estado siempre dispuesto a colaborar aportando su gran intuición y experiencia. En nuestras numerosas reuniones ha sabido guiarme de forma certera en la realización de este trabajo.

Juan Francisco Serra Cuñat

Madrid, Febrero de 2006

Capítulo 1

Medidas generalizadas de Entropía

1.1. Introducción

Diversos funcionales han sido propuestos en la literatura estadística como medidas de información siendo posible clasificarlos para su diferenciación en tres categorías: medidas paramétricas, no paramétricas y entropías.

- *Medidas paramétricas de información*: miden la cantidad de información aportada por los datos acerca de un parámetro desconocido θ y son funciones de θ, siendo la más conocida la medida de información de Fisher.

- *Medidas no paramétricas* (conocidas como divergencias): miden la "distancia" o afinidad entre dos distribuciones, o también la cantidad de información aportada por los datos a favor de una distribución F_1 y en contra de otra F_2, siendo la más conocidad la medida de Kullback-Leibler.

- *Medidas de Entropía*: miden la información contenida en una distribución, es decir, la incertidumbre acerca del resultado de un experimento, siendo las entropías de Shannon y de Rényi las medidas clásicas de este tipo.

Dadas las características de este trabajo nos centraremos exclusivamente en las medidas de entropía.

Las medidas de información (incertidumbre) conocidas como *entropías* tienen su origen en la Teoría de la Información, parte relativamente reciente de las matemáticas, pues

comienza a ser tratada con rigor a partir de la década de los cuarenta. Aunque posee un significado mucho más amplio (pensemos que el concepto de "información" es tan amplio que podría ser tratado desde un punto de vista puramente filosófico hasta un punto de vista estrictamente técnico), la Teoría de la Información se puede definir como el conjunto de problemas teóricos sobre transmisión de información a través de canales de comunicación incluyendo el estudio de medidas de información (incertidumbre) y de métodos óptimos de codificación de la información para su transmisión.

Los primeros estudios en esta dirección fueron realizados por Nyquist (1924), (1928) y Hartley (1928). Posteriormente en 1948 aparece el artículo de Claude Elwood Shannon *A Mathematical Theory of Communication,* publicado en *Bell System Technical Journal,* vol. 27 sobre las propiedades de las fuentes de información y de los canales de comunicación utilizados para la transmisión de información y que marca el comienzo de la Teoría de la Información como teoría matemática. Por la misma época e independientemente de Shannon, Wiener (1948) obtiene unos resultados similares, sin embargo, hay una diferencia de enfoques ya que en el modelo de Shannon, a diferencia del de Wiener, los mensajes son codificados antes de ser transmitidos. Ambos consideran como problema fundamental de la comunicación reconstruir exactamente o de la mejor forma posible el mensaje original a partir de la señal recibida.

Shannon establece las nociones de fuente de información, de canal de comunicación, de ruido en la transmisión, etc, y formula los teoremas fundamentales de la codificación que apoyan su teoría. Shannon vio que muchos de los problemas relacionados con la codificación, transmisión y decodificación de la información se podían tratar desde el punto de vista de una disciplina sistemática y matemática. La idea clave de la Teoría de la Información de Shannon es que la "información" puede medirse con una cantidad numérica (sobre la base de un modelo probabilístico) de forma que muchos problemas citados anteriormente pueden ser formulados en términos de esta medida de la cantidad de información.

Uno de los primeros problemas que tuvo que resolver Shannon fue el de definir el concepto de "información". Para Shannon este concepto va unido al de "incertidumbre": cuanto más incierto es un resultado, más información nos puede proporcionar cuando

se produce. Un experimento del cual sólo son posibles dos resultados, A y B, con la misma probabilidad de ocurrir contiene un *bit* de incertidumbre; y cuando el experimento se realiza, nos proporciona un *bit* de información. Shannon mide la información de un experimento a partir del promedio de las incertidumbres contenidas en cada uno de los resultados posibles del experimento. Este valor promedio, lo denomina *entropía*, nombre que le aconsejó John Van Neumann (a petición de Shannon) por la similitud de la expresión matemática de la medida de información de Shannon con la utilizada en la termodinámica estadística, ya que Shannon rehusaba utilizar el término "información" para su medida, pues consideraba que había sido utilizado en exceso (ver Tribus 1963).

El origen del concepto de entropía en Física, se encuentra en la Termodinámica (rama de la Física que estudia todos aquellos procesos en que interviene el calor). El concepto de entropía se inicia en la época de la evolución de la termodinámica clásica, es decir, cuando esta rama de la Física se ocupaba casi exclusivamente del estudio de las máquinas de vapor o, de forma más general, de las condiciones en las cuales se puede convertir el calor en trabajo, y no es un concepto probabilístico. En 1824 el físico francés Sadi Carnot, en su obra *Réflexions sur la puissance motrice du feu et les machines propres a développer cette puissance* propone el principio: "Una máquina térmica no puede funcionar sin el paso de calor de una fuente caliente a una fría", principio que en 1850, el físico alemán Clausius reformula diciendo: "el calor no puede pasar por sí mismo de un cuerpo frío a un cuerpo caliente", dando lugar a la noción de "entropía" (término que pone en circulación el propio Clausius) definida como una magnitud de estado del sitema considerado. La determinación de la entropía de una sustancia se reducía a medir cantidades de calor, es decir, a realizar medidas calorimétricas. A finales del siglo XIX se empieza a reconocer con Ludwig Boltzmann (creador, junto con J.W. Gibss, de la Mecánica Estadística, mediante la cual se puede dar un significado más profundo a las leyes y conceptos termodinámicos utilizando la concepción atómica) la naturaleza probabilística de la entropía; de hecho, Boltzmann (1896) fue el primero en dar un significado probabilístico a la entropía clásica. La entropía se determina en la mecánica estadística de una forma totalmente distinta a como se hace en la termodinámica clásica, pero ambos métodos dan por lo general el mismo resultado.

La conexión entre el concepto de entropía de la física y el de información es un asunto

todavía abierto, a pesar de las múltiples contribuciones al tema que se han producido. Las opciones van desde quien piensa que sólo hay una coincidencia en las fórmulas utilizadas, hasta quien opina que existe una identidad profunda, algo más que mera analogía, (véase por ejemplo el artículo de Weber, Depew, Dyke, Salthe, Schneider, Ulanowicz y Wicken, 1989). Pero la opinión más extendida actualmente es que conviene distinguir tres tipos de entropía: la que se utiliza en termodinámica clásica, la de la mecánica estadística y la informacional. Entre las dos primeras hay una estrecha y directa relación, mientras que la última es conceptualmente diferente y sólo se puede identificar con las anteriores en ciertos contextos físicos.

Una de las primeras aplicaciones directas de la Teoría de la Información fue su utilización en técnicas destinadas a mantener la seguridad en la trasmisión. El artículo de Shannon, *Communication Theory of Secrecy Systems* (1949), marca el comienzo de un estudio matemático basado en la Teoría de la Información y que ha dado lugar a técnicas muy sofisticadas para asegurar la confidencialidad en las transmisiones, garantizar la autenticidad del transmisor, etc.

Pasados los años cuarenta la literatura sobre la Teoría de la Información creció espectacularmente, y encontró aplicación en ingenierías, ciencias sociales, experimentales y biológicas; así ocurrió en economía, estadística, psicología, etc.

En Estadística, la utilización de las herramientas propias de la Teoría de la Información (medidas de información) para proporcionar métodos alternativos de estimación y contraste, a los clásicos, forman lo que hoy en día se conoce como *Teoría de la Información Estadística*.

Las medidas de entropía tratan de cuantificar la incertidumbre asociada a un experimento aleatorio. Pensemos por ejemplo en un experimento aleatorio A con dos posibles resultados con probabilidades $p_1,\ p_2\ \ (p_i \geq 0,\ i = 1, 2,\ p_1 + p_2 = 1)$ la incertidumbre acerca del posible resultado en caso de realizar el experimento depende de las probabilidades de los resultados, pues si se consideran los experimentos

$$A_1 \equiv \begin{pmatrix} a_1 & a_2 \\ 0.5 & 0.5 \end{pmatrix}, \qquad A_2 \equiv \begin{pmatrix} a_1 & a_2 \\ 0.999 & 0.001 \end{pmatrix}$$

el primer experimento aleatorio contiene más incertidumbre sobre el resultado que el segundo. Es lógico pensar que en A_2 el resultado a_1 ocurrirá "casi seguro". Las medidas de entropía asignan un valor numérico a cada distribución de probabilidad, materializando la idea intuitiva de mayor o menor incertidumbre. Por otra parte, las medidas de entropía pueden ser consideradas, también, como medidas para cuantificar el grado de homogeneidad con que la probabilidad se distribuye entre los distintos sucesos y por tanto como medidas de la "aleatoriedad" de una variable X, McEliece (1977).

¿Qué propiedades serían deseables desde un punto de vista intuitivo para una medida de incertidumbre?

Dado un experimento aleatorio A cuyos posibles resultados son $a_1, \ldots, a_n$ con probabilidades respectivas $p_1, \ldots, p_n$ $(p_i \geq 0,\ i = 1, \ldots, n,\ p_1 + \cdots + p_n = 1)$, una medida H de la incertidumbre contenida en A o proporcionada por A debería verificar:

1. Ser función de $p_1, p_2, \ldots, p_n$, por tanto se debe poder escribir como:

$$H(P) = H(p_1, p_2, \ldots, p_n)$$

2. Ser una función continua de $p_1, p_2, \ldots, p_n$, es decir, pequeños cambios en $p_1, p_2, \ldots, p_n$ deben producir pequeños cambios en H.

3. Debe conservar el valor numérico asociado a un experimento aleatorio, cuando se introduce en el experimento un resultado que no puede ocurrir

$$H_{n+1}(p_1, p_2, \ldots, p_n, 0) = H_n(p_1, p_2, \ldots, p_n)^1$$

4. Debe ser una función simétrica de sus argumentos

$$H(p_1, p_2, \ldots, p_n) = H(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(n)})$$

donde σ denota una permutación de $(1, \ldots, n)$.

[1]La notación $H_n(P) = H(p_1, \ldots, p_n)$ se utiliza solamente en aquellos casos en los que resulta imprescindible destacar el número de argumentos.

5. Debe tomar el valor cero cuando no existe incertidumbre, es decir,

$$H(p_1, p_2, \ldots, p_n) = 0 \text{ cuando } p_i = 1 \text{ para algún } i = 1, \ldots, n, \;\; p_j = 0 \;\; j \neq i$$

6. Debe tomar el valor máximo cuando todos los sucesos del experimento tienen la misma probabilidad de ocurrir (distribución uniforme)

$$p_1 = p_2 = \cdots = p_n = \frac{1}{n}$$

7. El valor máximo de $H_n(P)$ debe aumentar al crecer n (aumentar el número de los posibles resultados del experimento aleatorio)

8. Si A y B son dos experimentos aleatorios, independientes (el resultado de uno de ellos no influye sobre el otro) con posibles resultados $\{a_1, \ldots, a_n\}$, $\{b_1, \ldots, b_m\}$ y probabilidades $(p_1, \ldots, p_n)$, $(q_1, \ldots, q_m)$ respectivamente, el experimento compuesto $A \times B$ está formado por los sucesos $\{ a_i \cap b_j \,,\; i = 1, \ldots, n \,,\; j = 1, \ldots, m \}$ con probabilidades $(\, p_i q_j, \;\; i = 1, \ldots, n \,, \;\; j = 1, \ldots, m \,)$, entonces si

$$P * Q = (p_1 q_1, p_1 q_2, \ldots, p_1 q_m, \ldots, p_n q_1, \ldots, p_n q_m)$$

una buena propiedad sería que se verificase

$$H_{nm}(P * Q) = H_n(P) + H_m(Q) \qquad (\textit{Aditividad}).$$

1.2. Entropía de Shannon

Sea:

$$\triangle_n = \{P = (p_1, p_2, \ldots, p_n) \ : \ p_i \geq 0, \ i = 1, \ldots, n, \ \sum_{i=1}^{n} p_i = 1 \}$$

el conjunto formado por todas las distribuciones de probabilidad asociadas a una variable aleatoria discreta X que toma un número finito de valores $x_1, x_2, \ldots, x_n$; se denomina entropía de la variable aleatoria X o entropía de la distribución $P = (p_1, p_2, \ldots, p_n)$ a la expresión

$$H(X) = H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \, log p_i$$

Los logarítmos se pueden tomar con respecto a cualquier base que sea mayor que la unidad. En este trabajo, mientras no se diga lo contrario consideraremos base 2. La indeterminación $p_k \, log p_k$ con $p_k = 0$ se resuelve definiendo $p_k \, log p_k = 0$ si $p_k = 0$. Es decir, la función $f(x) = -x \, log x$ definida en $(0, \infty)$ se extiende por continuidad a $[0, \infty)$, definiendo

$$f(x) = \begin{cases} -x \, log x & \text{si } \ x > 0 \\ 0 & \text{si } \ x = 0 \end{cases}$$

Históricamente la entropía de Shannon fue la primera medida de información (incertidumbre), proporcionada por un experimento aleatorio, ya que la medida de Hartley (1928), único antecedente de la medida de Shannon, no es una medida de incertidumbre pues depende del número de resultados y no de la probabilidad de ocurrencia de los mismos.

Numerosas caracterizaciones se pueden encontrar en la literatura sobre la medida de Shannon (como solución de ecuación funcional, o via axiomática), se puede ver por ejemplo, Chaundy y McLeod (1960), Shannon (1948), Feinstein (1958), Aczél y Daróczy (1975) y Mathai y Rathie (1975).

La entropía de Shannon verifica un considerable número de interesantes propiedades (entre las que se encuentran todas las citadas anteriormente), ver Taneja (1990), de las que se han seleccionado las siguientes:

1. *No negatividad.* $H(P) \geq 0$. La igualdad se cumple si y sólo si $p_i = 1$ para algún i y $p_j = 0 \ \ (j \neq i)$.

2. *Continuidad.* $H(P)$ es una función continua de $p_1, \ldots, p_n$.

3. *Simetría.* $H(P)$ es una función simétrica de sus argumentos

$$H(p_1, \ldots, p_n) = H(p_{\sigma(1)}, \ldots, p_{\sigma(n)})$$

siendo σ una permutación de $(1, \ldots, n)$.

4. *Expansibilidad.*

$$H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n)$$

5. *Propiedad de la suma.*

$$H(P) = \sum_{i=1}^{n} f(p_i), \quad \text{donde} \quad f(p) = -p\, logp, \quad 0 \leq p \leq 1.$$

6. *Recursividad.*

$$H(p_1, \ldots, p_n) = H(p_1 + p_2,\, p_3, \ldots, p_n) + (p_1 + p_2)\, H\left(\frac{p_1}{p_1 + p_2},\, \frac{p_2}{p_1 + p_2}\right)$$

7. *Aditividad.*

$$H(P * Q) = H(P) + H(Q),$$

siendo $P * Q = (p_1 q_1 \ \ldots, p_1 q_m, p_2 q_1, \ldots, p_2 q_m, \ldots, p_n q_1, \ldots, p_n q_m)$,
$P \in \triangle_n, \ \ Q \in \triangle_m$.

8. *Agrupamiento.*

$$H(p_1, \ldots, p_n) = H(p_1 + \cdots + p_r \,,\, p_{r+1} + \cdots + p_n) + \left(\sum_{k=1}^{r} p_k\right) H\left(p_1 / \sum_{k=1}^{r} p_k, \ldots, p_r / \sum_{k=1}^{r} p_k\right)$$

$$+ \left(\sum_{k=r+1}^{n} p_k\right) H\left(p_{r+1} / \sum_{k=r+1}^{n} p_k, \ldots, p_n / \sum_{k=r+1}^{n} p_k\right)$$

9. *Valor máximo.* $H(P)$ alcanza el valor máximo con la distribución uniforme.

$$H(p_1, \ldots, p_n) \leq H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$$

con la igualdad para $p_i = \dfrac{1}{n}, \ \ \forall i = 1, \ldots, n$

10. *Propiedades relacionadas con la distribución uniforme.* Sea

$$\phi(n) = H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right), \quad n \geq 2, \quad n \in \mathbb{N}.$$

Entonces:

a) $\phi(n) \leq \phi(n+1)$.

b) $n\phi(n) \leq (n+1)\phi(n+1)$.

c) $\lim_{n\to\infty}\left[\phi(n+1) - \frac{n+1}{n}\phi(n)\right] = 0$

11. *Concavidad.* $H(P)$ es una función cóncava de P en $\triangle_n$.

12. *Schur-concavidad*

Definición 1.2.1. *Para todo P, $Q \in \triangle_n$ decimos que P está mayorizada por Q que denotamos $P \prec Q$ si*

a) $p_{(1)} \geq p_{(2)} \ldots \geq p_{(n)}$, $q_{(1)} \geq q_{(2)} \ldots \geq q_{(n)}$, con $\sum_{k=1}^{m} p_{(k)} \leq \sum_{k=1}^{m} q_{(k)}$, $1 \leq m \leq n$,

o

b) Existe una matriz doblemente estocástica (a_{kj}), $a_{kj} \geq 0$, $k, j = 1, \ldots, n$ tal que

$$p_{(k)} = \sum_{j=1}^{n} a_{kj} q_j \qquad k = 1, 2, \ldots, n$$

es decir p_k, $k = 1, \ldots, n$, es una media ponderada de las q_j, $j = 1, \ldots, n$.

Definición 1.2.2. *Una función $G : \triangle_n \to \mathbb{R}$ es Schur-cóncava en $\triangle_n$ si $P \prec Q$ implica $G(P) \geq G(Q)$.*

$H(P)$ es una función Schur-cóncava de P en $\triangle_n$.

13. Sea $\psi(p) = H(p,\ 1-p)$, $0 \leq p \leq 1$. Entonces

(i) $\psi(p) = \psi(1-p)$.

(ii) $\psi(1) = \psi(0)$.

(iii) $\psi(\frac{1}{2}) = 1$.

(iv) $\psi(p) + (1-p)\psi\left(\frac{q}{1-q}\right) = \psi(q) + \psi\left(\frac{p}{1-p}\right)$, $p, q \in [0, 1)$, $p + q \leq 1$.

14. Sea $p_{max} = max\{p_1, \ldots, p_n\}$. Entonces se verifica que

$$H(p_{max}\,,\ 1 - p_{max}) \leq H(P).$$

15. Diferencia entre dos entropías. Si

$$\sum_{i=1}^{n} |p_i - q_i| \leq \theta \leq \frac{1}{2}, \quad \text{entonces}$$

$$|H(P) - H(Q)| \leq -\theta\, log\frac{\theta}{n}, \qquad \forall\ P, Q \in \triangle_n.$$

1.3. Medidas generalizadas de entropía

Más de 30 medidas de entropía aparecen en la literatura de Teoría de la Información, generalizando la entropía de Shannon, entre las que cabe destacar las paramétricas, (introducidas por Rényi 1961), las trigonométricas (introducidas por Aczél y Daróczy 1963) y las ponderadas (introducidas por Belis y Guiasu 1968). Habitualmente, con el nombre de *entropías generalizadas* se denominan aquellas entropías dependientes de parámetros y tales que a partir de ellas, bien como valor particular de los mismos o como paso al límite, se obtiene la entropía de Shannon.

Hay dos métodos que son los que generalmente se utilizan en la caracterización de las entropías: uno consiste en proponer un conjunto de axiomas que debe verificar la entropía (via axiomática) y el otro utiliza ecuaciones funcionales cuya solución nos conduce a la entropía (ver como resumen por ejemplo, Aczél y Daróczy 1975 y Taneja 1979). Tres propiedades aparecen como más relevantes (juntas o individualmente) en la caracterización de las entropías que son: aditividad, recursividad y la propiedad de la suma.

Por último, en cuanto a la utilidad y ventajas que presentan cada una de ellas, hay que resaltar que están directamente relacionadas con el problema a tratar. En unos casos se primará la operatividad algebráica, en otros determinadas propiedades, etc.

1.3.1. Entropías paramétricas

- Entropía de orden r y de orden (r, s)

El primer intento para desarrollar una generalización de la entropía de Shannon fue llevado a cabo por Rényi (1961), el cual definió la entropía de orden r en los siguientes términos:

$$H_r(P) = \frac{1}{1-r}\, log\left(\sum_{i=1}^{n} p_i^r\right), \quad r \neq 1, \;\; r > 0. \tag{1.1}$$

para todo $P = (p_1, \ldots, p_n) \in \triangle_n$, siendo r un parámetro real. La entropía H_r contiene como caso límite la entropía de Shannon ya que se puede demostrar que

$$\lim_{r \to 1} H_r(P) = H(P)$$

siendo H(P) la entropía de Shannon.

En cuanto a sus aplicaciones, se pueden consultar entre otros Campbell (1965), Csiszár (1974), Kieffer (1979), Campbell (1985), Blumer y McEliece (1988).

Aczél y Daróczy (1963); Varma (1966), Kapur (1967) Rathie (1970) generalizan la entropía de orden r, siendo la estudiada por Aczél y Daróczy (1963) la que es conocida como entropía de orden (r, s) y cuya expresión es:

$$H_{r,s}(P) = \frac{1}{(s-r)}\, log \left(\frac{\sum_{i=1}^{n} p_i^r}{\sum_{i=1}^{n} p_i^s} \right), \quad r \neq s, \;\; r > 0, \;\; s > 0 \tag{1.2}$$

siendo r y s parámetros reales. En particular cuando $r = 1$ ó $s = 1$ la medida (1.2) se reduce a (1.1). También se puede demostrar que

$$\lim_{r \to s} H_{r,s}(P) = -\frac{\sum_{i=1}^{n} p_i^s \, log\, p_i}{\sum_{i=1}^{n} p_i^s}, \qquad s > 0$$

que se reduce a la entropía de Shannon para $s = 1$.

- **Entropía de grado s y grado (r, s)**

Por motivos operativos, parece más natural considerar la expresión $\sum_{i=1}^{n} p_i^r$ como medida de información en lugar de $log\left(\sum_{i=1}^{n} p_i^r\right)$. Por este motivo Havrda y Charvát (1967) proponen la siguiente entropía de grado s:

$$H^s(P) = (2^{1-s} - 1)^{-1} \left[\sum_{i=1}^{n} p_i^s - 1 \right], \quad s \neq 1, \;\; s > 0 \tag{1.3}$$

para todo $P = (p_1, \ldots, p_n) \in \triangle_n$. La entropía de grado s contiene como caso límite a la entropía de Shannon pues

$$\lim_{s \to 1} H^s(P) = H(P)$$

En el caso particular de $s = 2$, esta entropía conecta con el indice de Gini, el coeficiente de Bhattacharyya y la distancia Bayesiana, que se utilizan en otros campos además de la Teoría de la Información.

Caracterizaciones de esta entropía pueden consultarse en Havrda y Charvát (1967), Daróczy (1970).

Sharma y Taneja (1975, 1977) proponen una generalización de la entropía $\mathrm{H}^s(P)$ introduciendo dos parámetros, conocida como entropía de grado (r, s), cuya expresión es:

$$H^{r,s}(P) = (2^{1-r} - 2^{1-s})^{-1} \sum_{i=1}^{n} (p_i^r - p_i^s)\,, \quad r \neq s, \quad r > 0, \quad s > 0 \tag{1.4}$$

para todo $P = (p_1, \ldots, p_n) \in \triangle_n$, siendo r y s parámetros reales. En particular, cuando $r = 1$ ó $s = 1$ la medida anterior se reduce a la entropía de grado s y cuando $r \to s$

$$\lim_{r \to s} H^{r,s}(P) = -2^{r-1} \sum_{i=1}^{n} p_i^r \, log p_i\,, \quad r > 0$$

que se reduce a la entropía de Shannon, cuando $r = 1$.

- Entropía de clase t

Arimoto (1971) presentó otra generalización de la entropía de Shannon llamada entropía de clase t y que viene dada por

$${}_tH(P) = (2^{t-1} - 1)^{-1} \left[\left(\sum_{i=1}^{n} p_i^{1/t} \right)^t - 1 \right], \quad t \neq 1, \quad t > 0 \tag{1.5}$$

para todo $P = (p_1, \ldots, p_n) \in \triangle_n$. En este caso se verifica que

$$\lim_{t \to 1} {}_tH(P) = H(P).$$

- Entropías de orden 1 y grado s y orden r y grado s

Sharma y Mittal (1975) introducen y caracterizan dos entropías que denominan entropía de orden 1 y grado s y entropía de orden r y grado s dadas por las expresiones:

$$H_1^s(P) = (2^{1-s} - 1)^{-1} \left[exp_2 \left((s-1) \sum_{i=1}^{n} p_i \, log p_i \right) - 1 \right], \quad s \neq 1 \tag{1.6}$$

y

$$H_r^s(P) = (2^{1-s} - 1)^{-1} \left[\left(\sum_{i=1}^{n} p_i^r \right)^{\frac{s-1}{r-1}} - 1 \right], \quad r \neq 1, \quad s \neq 1, \quad r > 0 \tag{1.7}$$

La motivación de Sharma y Mittal fue generalizar las tres entropías, $H_r(P)$, $H^s(P)$ y ${}_tH(P)$. La relación entre ellas es la siguiente:

1. Cuando $r = s$, $H_r^s(P) = H_s^s(P) = H^s(P)$
2. Cuando $t = r^{-1} = 2 - s$, $H_r^s(P) = H_{\frac{1}{t}}^{2-t}(P) = {}_tH(P)$
3. $\lim_{s \to 1} H_r^s(P) = r H_r(P)$
4. $\lim_{r \to 1} H_r^s(P) = H_1^s(P)$
5. $\lim_{r \to 1} H_r(P) = \lim_{s \to 1} H^s(P) = \lim_{t \to 1} {}_tH(P) = \lim_{s \to 1} H_1^s(P) = H(P)$

Una relación detallada de las propiedades que verifican estas entropías puede consultarse en Taneja (1990). De entre estas merece la pena destacar la no negatividad, continuidad, simetría, seudoconcavidad, valor máximo, etc, siendo la propiedad de aditividad sustituida en la mayoría de los casos por la llamada *Seudoaditividad o No aditividad* dada por:

$$H(P * Q) = H(P) + H(Q) + C \, H(P) H(Q)$$

siendo C un valor numérico dependiente de la entropía considerada.

- Hypoentropías

Ferreri (1980) introduce un generalización de la entropía de Shannon llamada *Hypoentropía* dada por

$$H_\lambda(P) = \left(1 + \frac{1}{\lambda} \right) log(1 + \lambda) - \frac{1}{\lambda} \sum_{k=1}^{n} (1 + \lambda p_k) \, log(1 + \lambda p_k), \quad \lambda > 0.$$

Esta entropía contiene como caso límite la entropía de Shannon ya que

$$\lim_{\lambda\to\infty} H_\lambda(P) = H(P)$$

En Ferreri (1980) se pueden encontrar sus aplicaciones y sus propiedades entre las que se encuentran la recursividad y la propiedad de la suma y no verifican la propiedad aditiva.

1.3.2. Entropías trigonométricas

Introducidas por Aczél y Daróczy (1963), se agrupan en dos clases dependiendo de la función o funciones trigonométricas utilizadas. Por una parte, tenemos la entropía de Aczél y Daróczy (1963) dada por

$$\mathrm{S}(P) = \frac{1}{s} arctg \left\{ \sum_{i=1}^{n} p_i^r \, sen(s \, logp_i) / \sum_{i=1}^{n} p_i^r \, cos(s \, logp_i) \right\}, \quad s \neq 1, \quad s > 0, \quad r > 0$$

que se reduce a la entropía de Shannon cuando $r = 1$ y $s \to 1$. Por otro lado tenemos, las entropías de Sharma y Taneja (1977), Sant'anna y Taneja (1985), que utilizan solamente la función seno.

Sharma y Taneja (1977) proponen la siguiente entropía trigonométrica con dos parámetros:

$$\mathrm{S}_r^s(P) = -\frac{2^{r-1}}{sen\, s} \sum_{i=1}^{n} p_i^r \, sen(s \, logp_i), \quad r > 0, \quad s \neq k\pi, \quad k = 0, 1, \ldots$$

que para $r = 1$ se convierte en

$$\mathrm{S}_1^s(P) = -\frac{1}{sen\, s} \sum_{i=1}^{n} p_i \, sen(s \, logp_i), \quad s \neq k\pi, \quad k = 0, 1, \ldots$$

y cuando $s \to 0$

$$\lim_{s\to 1} \mathrm{S}_1^s(P) = H(P)$$

siendo $H(P)$ la entropía de Shannon.

En $\mathrm{S}_r^s(P)$ se ha utilizado la composición $sen(log(\cdot))$ pero también se pueden conseguir entropías generalizadas utilizando la composición $log(sen(\cdot))$. Por este motivo, Sant'anna y Taneja (1985) introducen y caracterizan las siguientes entropías trigonométricas dependientes de un parámetro:

1. $\mathrm{S}^s_{(1)}(P) = -\sum_{i=1}^{n} p_i \, log\left(\frac{sen\,(sp_i)}{2\,sen(s/2)}\right), \quad 0 < s < \pi$

2. $\mathrm{S}^s_{(2)}(P) = -\sum_{i=1}^{n} \left(\frac{sen\,(sp_i)}{2\,sen(s/2)}\right) log\left(\frac{sen\,(sp_i)}{2\,sen(s/2)}\right), \quad 0 < s < \pi$

3. $\mathrm{S}^s_{(3)}(P) = \sum_{i=1}^{n} \frac{sen\,(sp_i)}{2\,sen(s/2)}, \quad 0 < s < \pi$

Las dos primeras se reducen a la entropía de Shannon cuando $s \to 0$ mientras que la tercera, como caso excepcional, tiende a 1 cuando $s \to 0$, siendo comparable desde el punto de vista de las aplicaciones a la entropía de Shannon (Sant' anna y Taneja 1985).

1.3.3. Entropías con ponderaciones

La entropía fue introducida como medida cuantitativa de la información permitiendo tratar muchos de los problemas que constituyen la Teoría de la Información desde un punto de vista matemático, pero este resultado cuantitativo no agota todos los aspectos de la información.

En un sistema cibernético[2] (biológico o técnico) toda actividad está encaminada hacia la realización de un fin. El sistema debe disponer entonces de un criterio para poder diferenciar los sucesos. El criterio cibernético para la diferenciación cualitativa de los sucesos consiste en la importancia, la significación o la utilidad de la información que reportan respecto al fin. La aparición de un suceso elimina una doble "incertidumbre": una de orden cuantitativo relativa a la probabilidad de aparición y otra de orden cualitativo relativa a su utilidad para la realización del fin.

Basados en este planteamiento, Belis y Guiasu (1968) introducen y caracterizan (Guiasu 1977) la siguiente entropía con ponderaciones:

$$H(P;U) = -\sum_{i=1}^{n} p_i \, u_i \, log p_i$$

[2]Cibernética es la Ciencia que estudia comparativamente los sistemas de comunicación y regulación automática de los seres vivos con sistemas electrónicos y mecánicos semejantes a aquéllos.

donde $u_i \geq 0, \;\; i = 1, \ldots, n$ son los pesos o utilidades asociadas al suceso a_i con probabilidad p_i de ocurrir, y que permite diferenciar los sucesos según su importancia respecto al fin que se quiere alcanzar.

Basándose en la entropía de Belis y Guiasu, Picard (1979) presenta las siguientes generalizaciones:

$$H(P;V) = -\sum_{i=1}^{n} v_i \, log p_i \, / \sum_{i=1}^{n} v_i$$

$$H_r(P;V) = (1-r)^{-1} \, log \left(\sum_{i=1}^{n} p_i^{r-1} v_i \, / \sum_{i=1}^{n} v_i \right), \quad r \neq 1, \;\; r > 0$$

$$H_1^s(P;V) = (2^{1-s} - 1)^{-1} \left[exp_2 \left((s-1) \sum_{i=1}^{n} v_i \, log p_i \, / \sum_{i=1}^{n} v_i \right) \right]$$

$$H_r^s(P;V) = (2^{1-s} - 1)^{-1} \left[\left(\sum_{i=1}^{n} p_i^{r-1} v_i \, / \sum_{i=1}^{n} v_i \right)^{\frac{s-1}{r-1}} - 1 \right], \; r \neq 1, \; s \neq 1, \; r > 0, \; s > 0.$$

Otros trabajos sobre medidas de entropías con ponderaciones pueden verse en Emptoz, H. (1976), Gil, M. A., Pérez, R. y Gil, P. (1989), Pardo, L. (1986), Pardo, J.A. (1985, 1993, 1995), Pardo, J.A. y Pardo, M.C. (1995), etc.

Diversos funcionales se han propuesto en la literatura para recoger en una única expresión gran parte de las entropías citadas en este capítulo, ver por ejemplo, Salicrú, M.; Menendez, M. L., Morales, D. y Pardo, L. (1993) y Esteban, M. D.; Morales, D. (1995).

Por último cabe destacar también que en el artículo de Morales, D.; Pardo, L. y Vajda, I. (1996) se presenta un nuevo método de generar medidas de incertidumbre a partir de funciones schur-cóncavas.

1.4. Relación de entropías generalizadas

En la siguiente lista se recogen la mayoría de entropías generalizadas que aparecen en la literatura por orden cronologico con el nombre de sus respectivos autores, comenzando con la entropía de Shannon.

- *Shannon (1948)*

$$\Phi_1(P) = -\sum_{i=1}^{n} p_i log p_i$$

- *Rényi (1961)*

$$\Phi_2(P) = (1-r)^{-1} log\left(\sum_{i=1}^{n} p_i^r\right), \quad r \neq 1, \;\; r > 0$$

- *Aczél y Daróczy (1963)*

$$\Phi_3(P) = -\sum_{i=1}^{n} p_i^r log p_i \,/\, \sum_{i=1}^{n} p_i^r\,, \quad r > 0$$

$$\Phi_4(P) = (s-r)^{-1} log\left(\sum_{i=1}^{n} p_i^r \,/\, \sum_{i=1}^{n} p_i^s\right), \quad r \neq s, \;\; r > 0, \;\; s > 0$$

$$\Phi_5(P) = \frac{1}{s}\; arctg\left\{\sum_{i=1}^{n} p_i^r sen(s\, log p_i) \,/\, \sum_{i=1}^{n} p_i^r cos(s\, log p_i)\right\}, \;\; s \neq 1,\; s > 0,\; r > 0$$

- *Varma (1966)*

$$\Phi_6(P) = \frac{1}{m-r}\; log\left(\sum_{i=1}^{n} p_i^{r-m+1}\right), \quad m-1 < r < m, \;\; m \geq 1$$

$$\Phi_7(P) = \frac{1}{m(m-r)}\; log\left(\sum_{i=1}^{n} p_i^{r/m}\right), \quad 0 < r < m, \;\; m \geq 1$$

- *Kapur (1967)*

$$\Phi_8(P) = (1-t)^{-1}\; log\left(\sum_{i=1}^{n} p_i^{t+s-1} \,/\, \sum_{i=1}^{n} p_i^s\right), \quad t \neq 1, \;\; t > 0, \;\; s \geq 1$$

- *Havrda y Charvát (1967)*

$$\Phi_9(P) = (2^{1-s}-1)^{-1}\left[\sum_{i=1}^{n} p_i^s - 1\right], \quad s \neq 1, \;\; s > 0$$

- *Belis y Guiasu (1968)*

$$\Phi_{10}(P) = -\sum_{i=1}^{n} p_i u_i \, log p_i, \quad u_i > 0, \quad i = 1, \ldots, n$$

- *Rathie (1970)*

$$\Phi_{11}(P) = (1-r)^{-1} \, log\left(\sum_{i=1}^{n} p_i^{r+s_i-1} \Big/ \sum_{i=1}^{n} p_i^{s_i}\right), \;\; s_i \geq 0, \; i = 1, \ldots, n, \;\; r \neq 1, \; r > 0$$

- *Arimoto (1971)*

$$\Phi_{12}(P) = (2^{t-1}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^{1/t}\right)^t - 1\right], \quad t \neq 1, \;\; t > 0$$

- *Sharma y Mittal (1975)*

$$\Phi_{13}(P) = (2^{1-s}-1)^{-1}\left[exp_2\left((s-1)\sum_{i=1}^{n} p_i \, log p_i\right) - 1\right], \quad s \neq 1, \;\; s > 0$$

$$\Phi_{14}(P) = (2^{1-s}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^r\right)^{\frac{s-1}{r-1}} - 1\right], \;\; r \neq 1, \;\; s \neq 1, \;\; r > 0$$

- *Sharma y Taneja (1975; 1977)*

$$\Phi_{15}(P) = -2^{r-1}\sum_{i=1}^{n} p_i^r \, log p_i \, , \quad r > 0$$

$$\Phi_{16}(P) = (2^{1-r} - 2^{1-s})^{-1}\sum_{i=1}^{n} p_i^r - p_i^s, \quad r \neq s, \;\; r > 0, \;\; s > 0$$

$$\Phi_{17}(P) = -\frac{2^{r-1}}{sen \, s}\sum_{i=1}^{n} p_i^r \, sen(s \, log p_i), \quad r > 0, \;\; s \neq k\pi, \;\; k = 0, 1, \ldots$$

- *Picard (1979)*

$$\Phi_{18}(P) = -\sum_{i=1}^{n} v_i \, log p_i \, / \sum_{i=1}^{n} v_i$$

$$\Phi_{19}(P) = (1-r)^{-1} \, log \left(\sum_{i=1}^{n} p_i^{r-1} v_i \, / \sum_{i=1}^{n} v_i \right), \quad r \neq 1, \;\; r > 0$$

$$\Phi_{20}(P) = (2^{1-s} - 1)^{-1} \left[exp_2 \left((s-1) \sum_{i=1}^{n} v_i \, log p_i \, / \sum_{i=1}^{n} v_i \right) \right]$$

$$\Phi_{21}(P) = (2^{1-s} - 1)^{-1} \left[\left(\sum_{i=1}^{n} p_i^{r-1} v_i \, / \sum_{i=1}^{n} v_i \right)^{\frac{s-1}{r-1}} - 1 \right], \; r \neq 1, \; s \neq 1, \; r > 0, \; s > 0$$

- *Ferreri* (1980)

$$\Phi_{22}(P) = \left(1 + \frac{1}{\lambda} \right) \; log(1+\lambda) - \frac{1}{\lambda} \sum_{i=1}^{n} (1 + \lambda p_i) \, log(1 + \lambda p_i), \quad \lambda > 0$$

- *Sant'anna y Taneja (1985)*

$$\Phi_{23}(P) = -\sum_{i=1}^{n} p_i \, log \left(\frac{sen\,(s p_i)}{2 \, sen(s/2)} \right), \quad 0 < s < \pi$$

$$\Phi_{24}(P) = -\sum_{i=1}^{n} \left(\frac{sen\,(s p_i)}{2 \, sen(s/2)} \right) log \left(\frac{sen\,(s p_i)}{2 \, sen(s/2)} \right), \quad 0 < s < \pi$$

$$\Phi_{25}(P) = \sum_{i=1}^{n} \frac{sen\,(s p_i)}{2 \, sen(s/2)}, \quad 0 < s < \pi$$

- *Kapur (1988)*

$$\Phi_{26}(P) = -\sum_{i=1}^{n} log\, \Gamma(1 + p_i), \quad \text{siendo } \; \Gamma \; \text{ la función gamma.}$$

Capítulo 2

Optimización

2.1. Convexidad de conjuntos y funciones

2.1.1. Conjuntos convexos

Definición 2.1.1. *Conjunto convexo*

Dado un subconjunto S de $\mathbb{R}^n$ decimos que es convexo si para cada par de puntos $\bar{x},\ \bar{y} \in S$ y todo $\lambda \in [0,1]$ se verifica que

$$\bar{z} = \lambda\bar{x} + (1-\lambda)\bar{y} \in S$$

Propiedades

1. Sean $X_1, X_2, \ldots, X_n$ subconjuntos convexos de $\mathbb{R}^n$. Se verifica que

   $\bigcap_{i=1}^{n} X_i$ es un conjunto convexo.

2. La suma de n conjuntos convexos $X_1, X_2, \ldots, X_n$ de $\mathbb{R}^n$ definida como

   $$\sum_{i=1}^{n} X_i = \{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_n \in \mathbb{R}^n \ : \bar{x}_1 \in X_1\,,\ \bar{x}_2 \in X_2, \ldots, \bar{x}_n \in X_n\}$$

   es un conjunto convexo.

3. El producto de un conjunto convexo $X \in \mathbb{R}^n$ por un número real λ definido como

   $$\lambda X = \{\lambda\bar{x} : \ \bar{x} \in X\}$$

   es un conjunto convexo.

4. La combinación lineal de conjuntos convexos $X_1, \ldots, X_m \in \mathbb{R}^n$

$$X = \lambda_1 X_1 + \lambda_2 X_2 + \cdots \lambda_m X_m$$

es un conjunto convexo.

5. Sea A una transformación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$ definida

$$AC = \{A\bar{x} : \bar{x} \in C\} \qquad C \in \mathbb{R}^n$$

entonces AC es un conjunto convexo en $\mathbb{R}^m$ para cada conjunto convexo $C \in \mathbb{R}^n$

2.1.2. Funciones cóncavas y convexas

Sea M un subconjunto convexo y no vacío de $\mathbb{R}^n$ y f una función definida de M en $\mathbb{R}$. Entonces se dice que:

1. La función es convexa en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$ y para todo $\lambda \in [0,\ 1]$ se verifica que:

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) \leq \lambda f(\bar{x}) + (1-\lambda) f(\bar{y})$$

2. La función es cóncava en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$ y para todo $\lambda \in [0,\ 1]$ se verifica que:

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) \geq \lambda f(\bar{x}) + (1-\lambda) f(\bar{y})$$

3. La función es estrictamente convexa en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$ con $\bar{x} \neq \bar{y}$ y para todo $\lambda \in (0, 1)$ se verifica que:

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) < \lambda f(\bar{x}) + (1-\lambda) f(\bar{y})$$

4. La función es estrictamente cóncava en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$ con $\bar{x} \neq \bar{y}$ y para todo $\lambda \in (0, 1)$ se verifica que:

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) > \lambda f(\bar{x}) + (1-\lambda) f(\bar{y})$$

Propiedades de las funciones cóncavas y convexas

Sea M un subconjunto convexo de $\mathbb{R}^n$ y f una función definida de M en $\mathbb{R}$.

1. Si f es convexa en M entonces los conjuntos $\Lambda_\alpha = \{\bar{x} \in M \,/\, f(\bar{x}) \leq \alpha\}$ son convexos para todo $\alpha \in \mathbb{R}$

2. Si f es cóncava en M entonces los conjuntos $\Omega_\alpha = \{\bar{x} \in M \,/\, f(\bar{x}) \geq \alpha\}$ son convexos para todo $\alpha \in \mathbb{R}$

3. Si f es una función convexa en M, entonces $-f$ es cóncava.

4. Si f es una función estrictamente convexa en M, entonces $-f$ es una función estrictamente cóncava.

5. Si f es una función convexa en M y $\lambda \in \mathbb{R}$ entonces si $\lambda \geq 0$, la función λf es convexa y si $\lambda \leq 0$ la función λf es cóncava.

6. Si $\{f_i \,/\, i = 1, \ldots, m\}$ es una familia de funciones convexas en M entonces la función $f = \sum_{i=1}^{m} \alpha_i f_i$ con $\alpha_i \geq 0,\ i = 1, \ldots, m$ es una función convexa en M.

7. Si $f : \mathbb{R}^n \to \mathbb{R}$ es una función lineal entonces f es cóncava y convexa.

Condiciones para la convexidad de funciones diferenciables

Proposición 2.1.1.

Sea M un subconjunto abierto, no vacío y convexo de $\mathbb{R}^n$, y f una función diferenciable de M en $\mathbb{R}$. Se verifica que:

1. La función f es convexa en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$

$$f(\bar{y}) \geq f(\bar{x}) + \nabla f(\bar{x})(\bar{y} - \bar{x})$$

o bien

$$[\nabla f(\bar{y}) - \nabla f(\bar{x})](\bar{y} - \bar{x}) \geq 0$$

donde $\nabla f(\bar{x})$ denota el gradiente de f en $\bar{x}$.

2. La función f es estrictamente convexa en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$ con $\bar{x} \neq \bar{y}$

$$f(\bar{y}) > f(\bar{x}) + \nabla f(\bar{x})(\bar{y} - \bar{x})$$

o bien

$$[\nabla f(\bar{y}) - \nabla f(\bar{x})](\bar{y} - \bar{x}) > 0$$

3. La función f es cóncava en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$ se verifica

$$f(\bar{y}) \leq f(\bar{x}) + \nabla f(\bar{x})(\bar{y} - \bar{x})$$

o bien

$$[\nabla f(\bar{y}) - \nabla f(\bar{x})](\bar{y} - \bar{x}) \leq 0$$

4. La función f es estrictamente cóncava en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$ con $\bar{x} \neq \bar{y}$

$$f(\bar{y}) < f(\bar{x}) + \nabla f(\bar{x})(\bar{y} - \bar{x})$$

o bien

$$[\nabla f(\bar{y}) - \nabla f(\bar{x})](\bar{y} - \bar{x}) < 0$$

Definición 2.1.2. *Función de clase C^p*

Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}$ decimos que f es de clase C^p en A (abierto) si tiene derivadas parciales continuas en A hasta el orden p.

Proposición 2.1.2.

Sea M un subconjunto abierto, no vacío y convexo de $\mathbb{R}^n$ y f una función C^2, definida de M en $\mathbb{R}$, siendo $Hf(\bar{x})$ la matriz hessiana de f en $\bar{x}$. Entonces:

1. La función f es cóncava en M si y sólo si para todo $\bar{x} \in M$ se verifica que $\bar{y}' Hf(\bar{x})\bar{y} \leq 0$ para cualquier $\bar{y} \in \mathbb{R}^n$. Es decir, para todo $\bar{x} \in M$ la forma cuadrática con matriz asociada $Hf(\bar{x})$ es semidefinida negativa o definida negativa.

2. Si para todo $\bar{x} \in M$, se verifica que la forma cuadrática con matriz asociada $Hf(\bar{x})$ es definida negativa, la función f es estrictamente cóncava en M.

3. La función f es convexa en M si y sólo si, para todo $\bar{x} \in M$ se verifica que $\bar{y}' Hf(\bar{x})\bar{y} \geq 0$ para cualquier $\bar{y} \in \mathbb{R}^n$. Es decir, para todo $\bar{x} \in M$ la forma cuadrática con matriz asociada $Hf(\bar{x})$ es semidefinida positiva o definida positiva.

4. Si para todo $\bar{x} \in M$ la forma cuadrática con matriz asociada $Hf(\bar{x})$ es definida positiva, la función f es estrictamente convexa.

2.1.3. Funciones cuasicóncavas y seudocóncavas

Funciones cuasicóncavas

Sea M un subconjunto convexo y no vacío de $\mathbb{R}^n$ y f una función definida de M en $\mathbb{R}$. Entonces se dice que:

1. La función f es cuasicóncava en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M$ y para todo $\lambda \in [0,1]$ se verifica que:

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) \geq \min\{f(\bar{x}),\ f(\bar{y})\}$$

2. La función f es estrictamente cuasicóncava en M si y sólo si para cualesquiera $\bar{x},\ \bar{y} \in M,$ con $\bar{x} \neq \bar{y}$ y para todo $\lambda \in (0,1)$ se verifica que:

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) > \min\{f(\bar{x}),\ f(\bar{y})\}$$

Funciones seudocóncavas

Sea M un subconjunto convexo, abierto y no vacío de $\mathbb{R}^n$ y f una función definida de M en $\mathbb{R}$, diferenciable en M. Entonces se dice que:

1. La función f es seudocóncava en M si y sólo si se verifica una de las siguientes condiciones equivalentes.

   Para cualesquiera $\bar{x},\ \bar{y} \in M$ tales que $f(\bar{y}) > f(\bar{x})$ se tiene que $(\bar{y}-\bar{x})\nabla f(\bar{x}) > 0$.

   Para cualesquiera $\bar{x},\ \bar{y} \in M$ tales que $(\bar{y}-\bar{x})\nabla f(\bar{x}) \leq 0$ se tiene que $f(\bar{y}) \leq f(\bar{x})$.

2. La función f es estrictamente seudocóncava en M si y sólo si se verifica una de las siguientes condiciones equivalentes.

Para cualesquiera $\bar{x}, \bar{y} \in M$ con $\bar{x} \neq \bar{y}$ tales que $f(\bar{y}) \geq f(\bar{x})$ se tiene que $(\bar{y} - \bar{x})\nabla f(\bar{x}) > 0$.

Para cualesquiera $\bar{x}, \bar{y} \in M$, con $\bar{x} \neq \bar{y}$ tales que $(\bar{y} - \bar{x})\nabla f(\bar{x}) \leq 0$ se tiene que $f(\bar{y}) < f(\bar{x})$.

Proposición 2.1.3.

Sea M un subconjunto convexo y abierto de $\mathbb{R}^n$ y f una función de M en $\mathbb{R}$ cóncava y diferenciable en M. Entonces se verifica que f es seudocóncava.

Proposición 2.1.4.

Sea M un subconjunto de $\mathbb{R}^n$ convexo y f una función de M en $\mathbb{R}$ estrictamente seudocóncava. Entonces se verifica que f es estrictamente cuasicóncava.

Observaciones:

1. Otros autores bajo la denominación de función estrictamente cuasicóncava, enuncian conceptos distintos. Por otra parte las funciones estrictamente cuasicóncavas son también denominadas funciones X-cóncava, fuertemente cuasicóncavas, innominadas-cóncavas, etc.

2. Para más información sobre concavidad y concavidad débil ver Barbolla y Sanz (1995).

2.2. Programación matemática

La palabra "óptimo" como superlativo de "bueno" significa "sumamente bueno", "que no puede ser mejor". La optimización se puede considerar como la búsqueda de la mejor solución entre todas las posibles a un problema determinado. En la vida real practicamos habitualmente este ejercicio mental cuando elegimos entre diferentes opciones la más adecuada.

Una vez transcrito el problema considerado al lenguaje matemático, es preciso disponer de técnicas que nos permitan conocer si éste tiene o no solución y, en caso de tenerla, cuáles son su localización y naturaleza. Dada la diversidad de áreas en las que se plantean problemas de optimización, éstos tienen características muy diferentes. Por ello, también son necesarias técnicas distintas para poder abordarlos y resolverlos. La teoría que nos proporciona los resultados y herramientas precisos para estudiar este tipo de problemas es la Optimización Matemática.

El desarrollo de la Optimización Matemática no es reciente, ya que, aunque las aportaciones más importantes se produjeron en los años cuarenta y cincuenta del siglo XX, muchos de los resultados se conocían ya en el siglo XVIII. La *Programación Matemática* es una parte de la Teoría de la Optimización que incluye una gran variedad de problemas caracterizados fundamentalmente, con respecto a otros problemas de optimización, porque en ellos:

-Existe un único centro de decisión independiente. Lo que permite separar los problemas de Programación Matemática de los de la Teoría de Juegos.

-El tiempo no interviene como tal variable en la formulación del problema. Lo que nos permite diferenciar los problemas de Programación Matemática de los problemas de Optimización Dinámica.

Los problemas de Programación Matemática pueden definirse como los del cálculo del máximo o mínimo de una función de una o varias variables, cuando éstas se hallan sometidas a un conjunto de restricciones de distintos tipos. De acuerdo con esta definición, el objetivo de la Programación Matemática es el de calcular el mayor o el menor de los valores que puede tomar una función de los compatibles con las restricciones que pesan

sobre sus variables independientes.

Los programas matemáticos admiten la siguiente formulación general:

$$\left.\begin{array}{ll} Opt & f(x_1,\ldots,x_n) \\ s.a. & h_1(x_1,\ldots,x_n) = 0 \\ & \vdots \\ & h_m(x_1,\ldots,x_n) = 0 \\ & g_1(x_1,\ldots,x_n) \leq 0 \\ & \vdots \\ & g_k(x_1,\ldots,x_n) \leq 0 \\ & (x_1,\ldots,x_n) \in S \subset \mathbb{R}^n \end{array}\right\} \quad (P)$$

con f, h_i, $g_j : \mathbb{R}^n \to \mathbb{R}$ $i = 1,\ldots,m$, $j = 1,\ldots,k$ $(m < n)$.

Los elementos de un programa matemático son los siguientes:

$(x_1,\ldots,x_n)$	Variables de decisión o elección (tenemos que determinar sus valores).
$f(\bar{x})$	Función objetivo del problema.
Opt.	Optimizar la función f consiste en encontrar su máximo y su mínimo. Cuando únicamente se desea hallar el máximo, se escribe max, y en el caso de mínimo, min.
$h_i(\bar{x})$	Restricciones de igualdad que han de cumplir las posibles soluciones.
$g_j(\bar{x}) \leq 0$	Restricciones de desigualdad que han de cumplir las posibles soluciones.
$\bar{x} \in C$	Restricciones conjuntistas (variables enteras, dicotómicas, etc.).

Existe una gran variedad de programas matemáticos, con propiedades y métodos de solución diferentes. Los criterios de clasificación de dichos programas que habitualmente se utilizan son:

-*Tipo de restricciones* que intervienen en la formulación (sin restricciones, con restricciones de igualdad, etc.).

-*Tipo de funciones* que intervienen en la formulación, tanto la que define la función

objetivo como las que definen las restricciones (programas no lineales, lineales, etc.).

-*Número de variables y de restricciones* (pequeños, medianos, etc.).

-*Caraterísticas de convexidad y diferenciabilidad* de los conjuntos y funciones que intervienen en la formulación (programas diferenciables, convexos, etc.).

Definición 2.2.1. *Máximo y Mínimo globales*

Dado el programa matemático

$$\left.\begin{array}{l} Opt\ f(x_1,\ldots,x_n) \\ \bar{x}=(x_1,\ldots,x_n)\in B\subset\mathbb{R}^n \end{array}\right\}$$

1. Se dice que $\bar{x}^*\in B$ es máximo global del programa, si se verifica que

$$f(\bar{x})\leq f(\bar{x}^*),\ \text{ para todo }\ \bar{x}\in B$$

2. Se dice que $\bar{x}^*\in B$ es mínimo global del programa, si se verifica que

$$f(\bar{x})\geq f(\bar{x}^*),\ \text{ para todo }\ \bar{x}\in B$$

3. Se dice que $\bar{x}^*\in B$ es máximo global estricto del programa, si se verifica que

$$f(\bar{x})< f(\bar{x}^*),\ \text{ para todo }\ \bar{x}\in B\ \text{ con }\ \bar{x}\neq\bar{x}^*$$

4. Se dice que $\bar{x}^*\in B$ es mínimo global estricto del programa, si se verifica que

$$f(\bar{x})> f(\bar{x}^*),\ \text{ para todo }\ \bar{x}\in B\ \text{ con }\ \bar{x}\neq\bar{x}^*$$

Definición 2.2.2. *Máximo y Mínimo locales*

Dado el programa matemático

$$\left.\begin{array}{l} Opt\ f(x_1,\ldots,x_n) \\ \bar{x}=(x_1,\ldots,x_n)\in B\subset\mathbb{R}^n \end{array}\right\}$$

1. Se dice que $\bar{x}^*\in B$ es máximo local del programa si existe $r>0$ tal que

$$f(\bar{x})\leq f(\bar{x}^*)\ \text{ para todo }\ \bar{x}\in B(\bar{x}^*,r)\cap B$$

$B(\bar{x}^*, r)$ denota la bola abierta de centro $\bar{x}^*$ y radio r.

2. Se dice que $\bar{x}^* \in B$ es mínimo local del programa si existe $r > 0$ tal que

$$f(\bar{x}) \geq f(\bar{x}^*) \text{ para todo } \bar{x} \in B(\bar{x}^*, r) \cap B$$

3. Se dice que $\bar{x}^* \in B$ es máximo local estricto del programa si existe $r > 0$ tal que

$$f(\bar{x}) < f(\bar{x}^*) \text{ para todo } \bar{x} \in B(\bar{x}^*, r) \cap B \text{ con } \bar{x} \neq \bar{x}^*$$

4. Se dice que $\bar{x}^* \in B$ es mínimo local estricto del programa si existe $r > 0$ tal que

$$f(\bar{x}) > f(\bar{x}^*) \text{ para todo } \bar{x} \in B(\bar{x}^*, r) \cap B \text{ con } \bar{x} \neq \bar{x}^*$$

Definición 2.2.3. *Punto crítico*

Dada una función $f : S \subset \mathbb{R}^n \to \mathbb{R}$ diferenciable en S subconjunto abierto de $\mathbb{R}^n$, se dice que $\bar{x}^* \in S$ es un punto crítico de f cuando se verifica que $\nabla f(\bar{x}^*) = \bar{0}$.

Teorema 2.2.1. *Teorema de Weierstrass*

Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función continua en $A \subset \mathbb{R}^n$ y sea A un conjunto cerrado y acotado. Entonces existen $\bar{x}^*$, $\bar{x}^0 \in A$ tales que

$$f(\bar{x}^*) \leq f(\bar{x}) \quad \text{para todo} \quad \bar{x} \in A$$

$$f(\bar{x}^0) \geq f(\bar{x}) \quad \text{para todo} \quad \bar{x} \in A$$

es decir $\bar{x}^*$ es mínimo global de f en A y $\bar{x}^0$ es máximo global de f en A.

Teorema 2.2.2.

Si $f : A \subset \mathbb{R}^n \to \mathbb{R}$ es estrictamente convexa (cóncava) en A siendo A convexo y alcanza su valor mínimo (máximo) en un punto de A, este es único (por tanto global).

Definición 2.2.4. *Programa convexo*

Dado el programa matemático

$$\left.\begin{array}{ll} Opt & f(x_1,\ldots,x_n) \\ & (x_1,\ldots,x_n)\in B\subset\mathbb{R}^n \end{array}\right\}$$

Se dice que:

1. Es convexo para mínimo si B es convexo y f es una función convexa en B.

2. Es convexo para máximo si B es convexo y f es una función cóncava en B.

Teorema 2.2.3. *Teorema Fundamental de la Programación Convexa*

Dado el programa convexo

$$\left.\begin{array}{l} min\ f(x_1,\ldots,x_n) \\ s.a.\ \ (x_1,\ldots,x_n)\in B\subset\mathbb{R}^n \end{array}\right\}$$

se verifica que:

1. Si $\bar{x}^*\in B$ es un mínimo local, entonces $\bar{x}^*$ es un mínimo global.

2. El conjunto de todos los mínimos del programa es un conjunto convexo.

Para un problema de máximo se obtiene un resultado análogo sustituyendo el concepto de mínimo por el de máximo.

2.2.1. Optimización con restricciones de igualdad y desigualdad

Los programas matemáticos con restricciones de igualdad forman parte, junto con los programas sin restricciones, de la denominada " teoría clásica de la optimización", debido a que su solución teórica es conocida desde el matemático francés Lagrange (1736-1813).

La resolución de programas matemáticos con restricciones de desigualdad es mucho más reciente. En caso de programas lineales, la teoría y métodos de resolución de programas con este tipo de restricciones se conoce desde principios de los cincuenta, gracias a los trabajos del profesor estadounidense G. B. Dantzing. En programas con formulaciones no lineales, los métodos teóricos de resolución son conocidos a partir de los trabajos de los estadounidenses Kuhn y Tucker.

Optimización con restricciones de igualdad

La formulación general de un programa con restricciones de igualdad es

$$\left.\begin{array}{ll} Opt & f(x_1,\ldots,x_n) \\ s.a. & h_1(x_1,\ldots,x_n) = 0 \\ & \vdots \\ & h_m(x_1,\ldots,x_n) = 0 \\ & (x_1,\ldots,x_n) \in S \subset \mathbb{R}^n \end{array}\right\} \quad [\mathrm{I}]$$

con $m < n$ donde $f, h_j : \mathbb{R}^n \to \mathbb{R}, \;\; j = 1,\ldots,m,$

Teorema 2.2.4. *Teorema de Lagrange*

Sean $f, h_1, \ldots, h_m, \;\; m < n,$ funciones de clase C^1 en un subconjunto abierto $S \subseteq \mathbb{R}^n$ con valores en $\mathbb{R}$. Supongamos que $\bar{x}^* = (x_1^*, \ldots, x_n^*)$ es un óptimo local de f en el conjunto de soluciones factibles

$$B = \{(x_1,\ldots,x_n) \in S : \; h_j(x_1,\ldots,x_n) = 0, \; j = 1,\ldots,m\}$$

y supongamos también que los vectores $\nabla h_1(\bar{x}^*), \ldots, \nabla h_m(\bar{x}^*)$ son linealmente independientes. Entonces existen constantes $\lambda_1^*, \ldots, \lambda_m^*$ tales que

$$\nabla f(\bar{x}^*) + \sum_{j=1}^{m} \lambda_j^* \nabla h_j(\bar{x}^*) = \bar{0}. \qquad (1)$$

El Teorema de Lagrange presenta las condiciones necesarias de optimalidad local.

Se dice que en el punto $\bar{x} \in B$ (solución factible) se verifica la condición de regularidad o restricción de cualificación, si los vectores $\nabla h_1(\bar{x}), \ldots, \nabla h_m(\bar{x})$ son linealmente independientes. La condición de regularidad constituye una garantía de aplicabilidad del teorema de Lagrange.

Las soluciones factibles del programa [I] que verifican (1) se denominan *puntos estacionarios del programa.* Los m números reales $\lambda_1^*, \ldots, \lambda_m^*$ que se obtienen al resolver (1) se conocen como multiplicadores de Lagrange asociados a las m restricciones en el punto

$\bar{x}^*$.

Dado el programa [I] se denomina función Lagrangiana asociada al programa [I] (o simplemente Lagrangiano) a la función de $n+m$ variables L definida por

$$L(\bar{x},\, \bar{\lambda}) = f(\bar{x}) + \sum_{j=1}^{m} \lambda_j\, h_j(\bar{x})$$

con $\bar{x} = (x_1, \ldots, x_n)$ y $\bar{\lambda} = (\lambda_1, \ldots, \lambda_m)$.

En las hipótesis del teorema anterior, se verifica que todo punto crítico $(\bar{x}^*, \bar{\lambda}^*)$ de la función Lagrangiana asociada del programa [I], es un punto estacionario $\bar{x}^*$ del programa [I] con multiplicadores de Lagrange asociados $\bar{\lambda}^*$ como se puede comprobar fácilmente.

En la práctica, visto el resultado anterior, se suele construir la función Lagrangiana asociada al programa y se resuelve el sistema de $n+m$ ecuaciones con $n+m$ incógnitas

$$\begin{cases} \dfrac{\partial L(\bar{x},\, \bar{\lambda})}{\partial x_i} &= \dfrac{\partial f(\bar{x})}{\partial x_i} + \displaystyle\sum_{j=1}^{m} \lambda_j \frac{\partial h_j(\bar{x})}{\partial x_i} = 0 \qquad i = 1, \ldots, n \\ \dfrac{\partial L(\bar{x}, \bar{\lambda})}{\partial \lambda_j} &= h_j(\bar{x}) = 0 \qquad j = 1, \ldots, m \end{cases}$$

sus soluciones $(x_1^*, \ldots, x_n^*, \lambda_1^*, \ldots, \lambda_m^*)$ proporcionan, quedándose con las n primeras coordenadas $\bar{x}^* = (x_1^*, \ldots, x_n^*)$, candidatos a soluciones del programa. Si el programa tiene solución, ha de estar entre estos candidatos, por tanto se evalua la función f en cada uno de ellos y si se está maximizando el mayor de los valores obtenidos da la solución y si se está minimizando la solución la da el menor de ellos. Para dilucidar si existe solución se utilizan argumentos suplementarios basados en el teorema de Weierstrass o en propiedades de convexidad.

Si el **programa es convexo** las condiciones necesarias de Lagrange de optimalidad local son condiciones necesarias y suficientes de optimalidad **global**, (recordemos que si el programa no es convexo son solamente condiciones necesarias de optimalidad local)

Los programas convexos presentan, como se ve, enormes ventajas en el proceso de óptimización frente a otro tipo de programas, ya que todo punto que verifique las condiciones

de Lagrange se convierte en un óptimo global. Las condiciones suficientes de óptimo local en programas no convexos (de las cuales no se ha comentado nada ya que trabajaremos siempre con programas convexos) se pueden ver en Barbolla, R., Cerdá, E. y Sanz, P. (2000).

Optimización con restricciones de desigualdad

La formulación general de un programa con restricciones de desigualdad es

$$\left.\begin{array}{lll} Opt & f(x_1,\ldots,x_n) & \\ s.a. & h_1(x_1,\ldots,x_n) & \leq 0 \\ & \vdots & \\ & h_s(x_1,\ldots,x_n) & \leq 0 \\ & h_{s+1}(x_1,\ldots,x_n) & \geq 0 \\ & \vdots & \\ & h_m(x_1,\ldots,x_n) & \geq 0 \\ & (x_1,\ldots,x_n) \in & S \subset \mathbb{R}^n \end{array}\right\}$$

con $f,\ h_j : \mathbb{R}^n \to \mathbb{R} \quad j = 1,\ldots,m.$

El análisis del problema anterior, se puede reducir al estudio de

$$\left.\begin{array}{ll} min & f(x_1,\ldots,x_n) \\ s.a. & g_1(x_1,\ldots,x_n) \leq 0 \\ & \vdots \\ & g_m(x_1,\ldots,x_n) \leq 0 \\ & (x_1,\ldots,x_n) \in S \subset \mathbb{R}^n \end{array}\right\} \quad \text{[II]}$$

con $f,\ g_j : \mathbb{R}^n \to \mathbb{R},\ \ j = 1,\ldots,m,$ ya que $max\, f(x_1,\ldots,x_n)$ es equivalente a $min\ [-f(x_1,\ldots,x_n)]$ y las restricciones $h_l(x_1,\ldots,x_n) \geq 0,\ l = s+1,\ldots,m$ se pueden expresar como $[-h_l(x_1,\ldots,x_n)] \leq 0.$

Definición 2.2.5. *Restricción saturada*

Dada una solución factible $\bar{x}^*$ del problema [II], se dice que $\bar{x}^*$ satura la restricción i-ésima $g_i(\bar{x}^*) \leq 0$ si $g_i(\bar{x}^*) = 0$. Análogamente se dice que $\bar{x}^*$ no satura la restricción i-ésima si $g_i(\bar{x}^*) < 0$.

Teorema 2.2.5. *Teorema de Kuhn-Tucker*

Sean $f, g_1, \ldots, g_m$ funciones de clase C^1 en un subconjunto abierto $S \subseteq \mathbb{R}^n$ con valores en $\mathbb{R}$ y supongamos que $\bar{x}^* = (x_1^*, \ldots, x_n^*)$ es un mínimo local de f en el conjunto

$$B = \{\bar{x} \in S : g_j(\bar{x}) \leq 0,\ j = 1, \ldots, m\}$$

(conjunto factible) o sea, una solución local del problema [II]. Reordenando las funciones g_j si es necesario, podemos suponer que las restricciones de desigualdad que se saturan en $\bar{x}^*$, son $g_1(\bar{x}^*) = 0, \ldots, g_r(\bar{x}^*) = 0$, con $r \leq m$. Pues bien, si los vectores $\nabla g_1(\bar{x}^*), \ldots, \nabla g_r(\bar{x}^*)$ son linealmente independientes, entonces existen constantes $\lambda_1^*, \ldots, \lambda_m^*$ tales que

$$\nabla f(\bar{x}^*) + \sum_{j=1}^{m} \lambda_j^* \nabla g_j(\bar{x}^*) = \bar{0}$$

$$\begin{aligned} \lambda_j^* g_j(\bar{x}^*) &= 0 \quad \text{para } j = 1, \ldots, m \\ \lambda_j^* &\geq 0 \qquad \text{para } j = 1, \ldots, m \\ g_j(\bar{x}^*) &\leq 0 \qquad \text{para } j = 1, \ldots, m \end{aligned}$$

Este teorema recoge las condiciones necesarias de optimalidad local. A los escalares $\lambda_j^*,\ \ j = 1, \ldots, m$ se les denomina multiplicadores de Kuhn-Tucker asociados a las m restriciones en el punto $\bar{x}^*$. Si el programa se plantea en los términos

$$\left.\begin{array}{lll} max & f(x_1, \ldots, x_n) & \\ s.a. & g_1(x_1, \ldots, x_n) & \leq\ 0 \\ & \vdots & \\ & g_m(x_1, \ldots, x_n) & \leq\ 0 \end{array}\right\}$$

las condiciones necesarias de Kuhn-Tucker se expresan como sigue:

$$\nabla f(\bar{x}^*) + \sum_{j=1}^{m} \lambda_j^* \nabla g_j(\bar{x}^*) = \bar{0}$$

$$\lambda_j^* g_j(\bar{x}^*) = 0 \quad \text{para} \;\; j = 1, \ldots, m$$

$$\lambda_j^* \leq 0 \qquad \text{para} \;\; j = 1, \ldots, m$$

$$g_j(\bar{x}^*) \leq 0 \qquad \text{para} \;\; j = 1, \ldots, m$$

La *condición de regularidad* es la independencia lineal de los vectores $\nabla g_1(\bar{x}^*), \ldots, \nabla g_r(\bar{x}^*)$. (gradientes de las restricciones saturadas) que constituye una garantía de aplicabilidad del Teorema de Kuhn-Tucker.

En la práctica se actúa de forma similar a la vista en el caso anterior (restricciones de igualdad): se construye la función Lagrangiana

$$L(\bar{x}, \bar{\lambda}) = f(\bar{x}) + \sum_{j=1}^{m} \lambda_j \, g_j(\bar{x})$$

y se resuelve el sistema de *condiciones de Kuhn-Tucker*

$$\begin{cases} \dfrac{\partial L(\bar{x}, \bar{\lambda})}{\partial x_i} = \dfrac{\partial f(\bar{x})}{\partial x_i} + \displaystyle\sum_{j=1}^{m} \lambda_j \frac{\partial g_j(\bar{x})}{\partial x_i} = 0, \qquad i = 1, \ldots, n \\ g_j(\bar{x}) \leq 0 \qquad j = 1, \ldots, m \\ \lambda_j \, g_j(\bar{x}) = 0 \qquad j = 1, \ldots, m \\ \lambda_j \geq 0 \quad \text{para minimizar;} \quad \lambda_j \leq 0 \;\; \text{para maximizar} \end{cases}$$

sus soluciones $(\bar{x}_1^*, \ldots, x_n^*, \lambda_1^*, \ldots, \lambda_m^*)$ proporcionan, quedándose con las n primeras coordenadas $(x_1^*, \ldots, x_n^*)$, candidatos a soluciones del programa. Si éste tiene solución, ha de estar entre estos candidatos, por tanto se evalúa la función f en cada uno de ellos y si se está maximizando el mayor de los valores obtenidos, da la solución y si se está minimizando la solución la da el menor de ellos.

Existe una gran similitud entre los multiplicadores de Kuhn-Tucker asociados a programas con restricciones de desigualdad y los multiplicadores de Lagrange asociados a programas con restricciones de igualdad. Básicamente, la diferencia entre ambos consiste en que los multiplicadores de Lagrange asociados a programas con restricciones

de igualdad pueden tomar cualquier signo, mientras que los multiplicadores de Kuhn-Tucker asociados a programas con restricciones de desigualdad deben ser no positivos o no negativos, según la formulación del problema.

Los programas de minimización y maximización pueden formularse también con las restricciones en forma $g_j(\bar{x}) \geq 0, \quad j = 1, \ldots, m$. Esta modificación en la formulación del programa afecta al signo de los escalares $\lambda_j, \quad j = 1, \ldots, m$. En concreto, para las cuatro posibles formulaciones los cambios se recogen en el siguiente cuadro

	min	max
$\bar{g}(\bar{x}) \leq 0$	$\bar{\lambda} \geq 0$	$\bar{\lambda} \leq \bar{0}$
$\bar{g}(\bar{x}) \geq 0$	$\bar{\lambda} \leq 0$	$\bar{\lambda} \geq \bar{0}$

Las condiciones de Kuhn y Tucker constituyen condiciones necesarias de optimalidad local y son solamente aplicables a programas diferenciables, es decir, a programas en los que las funciones que intervienen en su definición (objetivo y restricciones) son funciones diferenciables.

Si el **programa es convexo** las condiciones de Kuhn-Tucker de optimalidad local son condiciones necesarias y suficientes de optimalidad **global**. (Si el programa no es convexo son solamente condiciones necesarias de optimalidad local).

Las condiciones suficientes de óptimo local para programas no convexos se pueden ver en Barbolla, R., Cerdá, E. y Sanz, P. (2000).

El problema general de optimización

El problema general de optimización es aquel que incluye a la vez resticciones de igualdad y restricciones de desigualdad (ver Fernández C., Hernández, F. J.,Vegas J.M. 2002). La formulación general de un programa con restricciones de igualdad y desigualdad es:

$$\left.\begin{array}{ll} Opt & f(x_1,\ldots,x_n) \\ s.a. & h_1(x_1,\ldots,x_n) = 0 \\ & \vdots \\ & h_s(x_1,\ldots,x_n) = 0 \\ & g_1(x_1,\ldots,x_n) \leq 0 \\ & \vdots \\ & g_m(x_1,\ldots,x_n) \leq 0 \\ & (x_1,\ldots,x_n) \in S \subset \mathbb{R}^n \end{array}\right\} \quad \text{[III]}$$

con $f,\ h_k,\ g_j : \mathbb{R}^n \to \mathbb{R}, \quad k = 1,\ldots,s, \ \ j = 1,\ldots,m.$

El número de restricciones de igualdad tiene que ser menor que el de variables de decisión $(s < n)$. De los teoremas de Lagrange y Kuhn-Tucker se deducen las condiciones necesarias que debe cumplir un punto $\bar{x}^* \in S$ para que sea solución óptima de [III] estas condiciones vienen dadas por el siguiente teorema

Teorema 2.2.6.

Sean $f, g_1,\ldots,g_m,\ h_1,\ldots,h_s,\ (s < n)$, funciones de clase C^1 en un subconjunto abierto $S \subseteq \mathbb{R}^n$ con valores en $\mathbb{R}$. Supongamos que $\bar{x}^* = (x_1^*, x_2^*,\ldots,x_n^*)$ es un óptimo local de f en el conjunto

$$B = \{\bar{x} \in S \ :\ g_j(\bar{x}) \leq 0,\ j = 1,\ldots m;\ \ h_k(\bar{x}) = 0,\ k = 1,\ldots,s\}$$

o sea, una solución local del problema [III]. Reordenando las funciones g_j si es necesario, podemos suponer que las restricciones de desigualdad que se saturan en $\bar{x}^*$ son $g_1(\bar{x}^*) = 0,\ldots,g_r(\bar{x}^*) = 0,$ con $r \leq m$. Pues bien, si los vectores

$$\{\nabla g_1(\bar{x}^*),\ldots,\nabla g_r(\bar{x}^*),\ \nabla h_1(\bar{x}^*),\ldots,\nabla h_s(\bar{x}^*)\}$$

son linealmente independientes, entonces existen constantes $\lambda_1^*, \ldots, \lambda_m^*$ y $\mu_1^*, \ldots, \mu_s^*$ tales que

$$\begin{cases} \nabla f(\bar{x}^*) + \sum_{j=1}^{m} \lambda_j^* \nabla g_j(\bar{x}^*) + \sum_{k=1}^{s} \mu_k^* \nabla h_k(\bar{x}^*) = \bar{0} \\ \lambda_j^* g_j(\bar{x}^*) = 0 \text{ para } j = 1, \ldots, m \\ \\ g_j(\bar{x}^*) \leq 0 \text{ para } j = 1, \ldots, m \\ \\ \lambda_j^* \geq 0 \text{ si } \bar{x}^* \text{ es un mínimo, } \lambda_j^* \leq 0 \text{ si } \bar{x}^* \text{ es un máximo.} \end{cases}$$

La *condición de regularidad* es la independencia lineal de los vectores

$$\{\nabla g_1(\bar{x}^*), \ldots, \nabla g_r(\bar{x}^*),\ \nabla h_1(\bar{x}^*), \ldots, \nabla h_s(\bar{x}^*)\}$$

(gradientes de las restricciones saturadas y gradientes de las restricciones de igualdad) que constituye una garantía de aplicabilidad del teorema.

Si el **programa es convexo** (como será en nuestro caso) las condiciones de optimalidad local anteriores son condiciones necesarias y suficientes de optimalidad global.

En la práctica se construye la función Lagrangiana

$$L(\bar{x}, \bar{\lambda}, \bar{\mu}) = f(\bar{x}) + \sum_{j=1}^{m} \lambda_j\, g_j(\bar{x}) + \sum_{k=1}^{s} \mu_k h_k(\bar{x})$$

y se resuelve el sistema de *condiciones de Kuhn-Tucker*

$$\begin{cases} \dfrac{\partial L}{\partial x_i} = \dfrac{\partial f(\bar{x})}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \dfrac{\partial g_j(\bar{x})}{\partial x_i} + \sum_{k=1}^{s} \mu_k \dfrac{\partial h_k(\bar{x})}{\partial x_i} = 0, \qquad i = 1, \ldots, n \\ h_k(\bar{x}) = 0 \qquad k = 1, \ldots, s \\ g_j(\bar{x}) \leq 0 \qquad j = 1, \ldots, m \\ \lambda_j\, g_j(\bar{x}) = 0, \qquad j = 1, \ldots, m \\ \lambda_j \geq 0 \text{ para minimizar; } \lambda_j \leq 0 \text{ para maximizar} \end{cases}$$

Sus soluciones $(\bar{x}_1^*, \ldots, x_n^*, \lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_s^*)$ proporcionan, quedándose con las n primeras coordenadas $(x_1^*, \ldots, x_n^*)$, candidatos a soluciones del programa y se procede como en los casos anteriores.

Capítulo 3

Presentación, análisis y resolución del problema

3.1. Presentación del problema

- Sea $\triangle_k = \{\, \theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k : \ \theta_j \geq 0, \ j = 1, \ldots, k, \ \sum_{j=1}^{k} \theta_j = 1 \,\}, \ \ k \geq 2.$

- Sean $c_1, c_2, \ldots, c_k,$ números reales conocidos tales que $c_i \geq 0, \ i = 1, \ldots, k$ y $\sum_{i=1}^{k} c_i = 1.$

- Sea $\omega = (\omega_1, \omega_2, \ldots, \omega_r)$ la siguiente transformación lineal de θ para cada r fijo, entre k y k^2 :

$$\omega_i = \theta_1 \sum_{l \in L_{i1}} c_l + \theta_2 \sum_{l \in L_{i2}} c_l + \cdots + \theta_k \sum_{l \in L_{ik}} c_l; \qquad i = 1, \ldots, r$$

donde los conjuntos L_{ij} se definen de la siguiente forma:

Para j fijo $(1, \ldots, k)$ los elementos L_{ij} del conjunto $\{L_{ij}\}_{i=1}^{r}$ verifican:

$$L_{i_1 j} \cap L_{i_2 j} = \emptyset, \ \ i_1 \neq i_2 = 1, \ldots, r$$

$$\bigcup_{i=1}^{r} L_{ij} = \{1, 2, \ldots, k\}$$

Por tanto, para r fijo, $k \leq r \leq k^2$, ω define una distribución de probabilidad finita

cuya variable aleatoria discreta asociada toma un número r de valores. Es decir:

$$\omega_i \geq 0, \quad i = 1, \ldots, r, \quad \sum_{i=1}^{r} \omega_i = 1.$$

Conceptualmente tenemos un experimento con k posibles resultados $(R_1, \ldots, R_k)$ con distribución de probabilidad $(\theta_1, \ldots, \theta_k)$. Si la transformación lineal redujese la dimensión: $(\omega_1, \ldots, \omega_r)$, $r < k$, el nuevo experimento observado consistiría en r resultados $(R'_1, \ldots, R'_r)$ y nunca podría contener como caso particular el experimento $(R_1, \ldots, R_k)$. Esta es una de las razones de considerar $r \geq k$; posteriormente veremos otras razones de tipo operacional que también justifican esta consideración.

Las probabilidades $\omega_i = \omega_i(\theta)$ pueden también definirse a partir del producto Kronecker $\theta * c$. Dichas probabilidades $\omega_i(\theta)$, $i = 1, \ldots, r$ son sumas de probabilidades de $\theta * c$, es decir:

$$\omega_i(\theta) = \sum_{j_1\,,\,j_2} \theta_{j_1}\, c_{j_2}$$

Otra forma de definir ω es mediante la siguiente ecuación matricial:

$$\omega \;=\; A\,\theta$$

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_r \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rk} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{pmatrix}$$

donde $A = (a_{ij})$, $\quad a_{ij} \geq 0\,,\;\; i = 1, \ldots, r\,,\;\; j = 1, \ldots, k$

siendo:

$$a_{ij} = \sum_{l \in L_{ij}} c_l \quad i = 1, \ldots, r\,,\;\; j = 1, \ldots, k$$

Claramente para cualquier $j = 1, \ldots, k$ se tiene que

$$\sum_{i=1}^{r} a_{ij} = \sum_{i=1}^{r} \sum_{l \in L_{ij}} c_l = \sum_{l=1}^{k} c_l = 1$$

la última igualdad se debe a que $\bigcup_{i=1}^{r} L_{ij} = \{1, 2, \ldots, k\}$ para todo $j = 1, \ldots, k$. Por lo tanto, la matriz A es una matriz estocástica.

Para $c = (c_1, \ldots, c_k)$ fijo, denotamos por Ω_r^* al conjunto:

$$\Omega_r^* = \{\, \omega = (\omega_1(\theta), \ldots, \omega_r(\theta)) \in \mathbb{R}^r, \;\; \omega = A\theta, \;\; \theta \in \triangle_k\}$$

Para todo c y r se verifica que $\omega_i(\theta) \geq 0, \;\; i = 1, \ldots, r$ y $\sum_{i=1}^{r} \omega_i(\theta) = 1$. Claramente, $\Omega_r^* \subseteq \Omega_r$. siendo

$$\Omega_r = \{\, \omega = (\omega_1, \ldots, \omega_r) \in \mathbb{R}^r : \; \omega_j \geq 0, \; j = 1, \ldots, r, \;\; \sum_{j=1}^{r} \omega_j = 1 \,\}$$

En el caso de que $\omega_i(\theta) = 0, \;\; \forall\, \theta \in \triangle_k$ para al menos un i, lo cual ocurre cuando $\sum_{j=1}^{k} a_{ij} = 0 \Leftrightarrow a_{ij} = 0 \;\; \forall j,$ la dimensión de Ω_r^* sería menor que r por lo cual el problema inicial se definiría con una nueva matriz A obtenida de la anterior excluyendo la fila(s) de ceros correspondiente(s).

En el caso particular de $r = k,$ la matriz A es cuadrada, si además A es no singular, ésta define un transformación lineal biyectiva de $\mathbb{R}^k$ en $\mathbb{R}^k$. También, en este caso (matriz A cuadrada) si todos los conjuntos L_{ij} contienen un único elemento y $L_{ij_1} \cap L_{ij_2} = \emptyset, \;\; j_1 \neq j_2 = 1, \ldots, k \;\; \forall i,$ entonces la matriz A es doblemente estocástica y $\omega \prec \theta, \; \omega$ está mayorizada por θ (Definición 1.2.1).

- Sea $\mathcal{H}$ el conjunto de entropías, definidas sobre Ω_r cuyos elementos verifican las propiedades (ver sección 1.2.) de continuidad, simetría, siendo además funciones que alcanzan su valor máximo con la distribución uniforme (propiedades que verifican las medidas de entropía presentadas en el capítulo I).

Se quiere estudiar el comportamiento de $H \in \mathcal{H}$ como función de $\theta, \; H(\omega(\theta)) = H(A\theta)$ más concretamente, caracterizar el valor o valores de θ que maximizan dicha entropía. Fijada $H \in \mathcal{H}$ hay que resolver el siguiente programa matemático

$$\left.\begin{array}{l} \max H(A\theta) \\ s.a. \\ \theta_j \geq 0 \;\; j = 1, \ldots, k \\ \sum_{j=1}^{k} \theta_j = 1 \end{array}\right\} \qquad [\mathrm{I}]$$

3.2. Método alternativo

Para calcular analíticamente la solución del programa anterior hay que aplicar el teorema de Kuhn-Tucker que exige la diferenciabilidad de H y resolver sistemas de ecuaciones no lineales bastante complejos (formados a partir de las condiciones de Kuhn-Tucker), lo que supone en la mayoría de los casos la necesidad de utilizar métodos numéricos que nos proporcionan soluciones aproximadas. Nosotros proponemos a continuación un método para la obtención de una solución aproximada del programa anterior, solución que presenta dos cualidades fundamentales:

1. Se puede considerar bajo determinadas condiciones, como solución del programa [I] independientemente de la entropía considerada ya que, entonces, el error cometido es despreciable.

2. Puede servir como punto inicial para los métodos de optimización denominados "métodos de busqueda directa" que se caracterizan por la no utilización explícita de las derivadas de la función objetivo en las técnicas de optimización y que permiten también como caso especial, calcular la solución del programa [I] cuando $H \in \mathcal{H}$ no es diferenciable.

El método consiste en resolver el sistema (cuando sea compatible)

$$\omega^{(0)} = A\theta \qquad (*)$$

con $\omega^{(0)} = (1/r, \ldots, 1/r)$, es decir, encontrar $\theta \in \triangle_k$ que se transforma en la distribución uniforme, que es justamente la distribución en la que cualquier $H \in \mathcal{H}$ alcanza el valor máximo absoluto. Si el sistema es incompatible se busca una "seudosolución" conocida en la literatura como *solución mínimo cuadrática del sistema* $(*)$, mediante la resolución del siguiente programa matemático

$$\left.\begin{array}{l} \text{mín } ||\omega^{(0)} - A\theta|| \\ s.a. \\ \theta_j \geq 0, \quad j = 1, \ldots, k \\ \displaystyle\sum_{j=1}^{k} \theta_j = 1 \end{array}\right\} \qquad \text{[II]}$$

con $||\cdot||$ la norma euclídea.

Entre las ventajas que aporta este método cabe destacar:

a) A la solución del programa [II] se llega mediante la resolución de sistemas de ecuaciones lineales.

b) Para algunas entropías de la familia $\mathcal{H}$ como por ejemplo: Rényi de orden 2 y Havrda y Charvát de grado 2, la solución del programa [I] coincide con la del programa [II].

c) Respeta la idea intuitiva de que la distribución $\theta \in \triangle_k$ solución de [II], o es la que se transforma en la distribución uniforme $\omega^{(0)}$, o es la que se transforma en la más parecida (próxima en norma ecuclídea a $\omega^{(0)}$).

3.2.1. Características de las soluciones de [I] y [II]

1. Sea θ_H^* la solución del programa [I] y $\omega_H^* = A\theta_H^*$, sea θ^* la solución del programa [II] y $\omega^* = A\theta^*$, es decir, ω^* es el punto más próximo a $\omega^{(0)}$ dentro del conjunto Ω_r^*; $\omega_H^* = \omega^*$ cuando la proyección del vector gradiente $\nabla H(\omega^*)$ sobre el plano determinado por los puntos ω^*, ω_H^* y $\omega^{(0)}$ tiene la misma dirección y sentido que el vector $\bar{v} = \overline{\omega^*\omega^{(0)}}$. Esta condición se da obviamente para entropías $H \in \mathcal{H}$ con conjuntos de nivel definidos por puntos ω equidistantes de $\omega^{(0)}$ como son Rényi de orden 2 y Havrda y Charvát de grado 2.

2. Una vez fijada la variedad lineal $\omega = A\theta$ cada medida de entropía H localizará el punto óptimo sobre ella, ω_H^*, a partir del conjunto de nivel tangente a dicha variedad. Por contra ω^* no depende de la medida H elegida.

3. Para cualquier $H \in \mathcal{H}$ se tiene

$$H(\omega_H^*) - H(\omega^*) < H(\omega^{(0)}) - H(\omega^*)$$

la anterior desigualdad nos proporciona una primera valoración del error cometido.

Dadas las características de los programas [I] y [II] se podría pensar en la equivalencia entre ambos programas, el siguiente contraejemplo, pone de manifiesto que de forma general no existe tal equivalencia.

Contraejemplo. Sea

$$\omega = \begin{pmatrix} 0.2 & 0 \\ 0.8 & 0.2 \\ 0 & 0{,}8 \end{pmatrix} \theta$$

con $\theta \in \triangle_2$ y $\omega \in \Omega_3^* \subset \Omega_3$ siendo

$$\Omega_3^* = \{\, \omega = (\omega_1(\theta), \omega_2(\theta), \omega_3(\theta)) \in \Omega_3 / \;\; \omega = A\theta, \;\; \theta \in \triangle_2\}$$

que se puede expresar como:

$$\Omega_3^* = \{\omega \in \Omega_3 / \, \omega_2 = 3\omega_1 + 0.2\,, \;\; \omega_3 = 1 - \omega_1 - \omega_2\}$$

Se puede comprobar que la solución del programa [I] (que se ha obtenido utilizando un programa de cálculo simbolico) con la entropía de Shannon es $\theta_{Sh}^* = (0.55, 0.45)$, que se transforma por $\omega = A\theta$ en $\omega_{Sh}^* = (0.11, 0.53, 0.36)$ como el punto de la variedad lineal Ω_3^* de máxima entropía.

Por otra parte la solución del programa [II] es $\theta^* = (0.5, 0.5)$ que se transforma por $\omega = A\theta$ en el punto $\omega^* = (0.1, 0.5, 0.4)$ como el más próximo de los pertenecientes a la variedad lineal Ω_3^* a $\omega^{(0)} = (1/3, 1/3, 1/3)$, por tanto, los dos programas no son equivalentes. Los valores de la entropía de Shannon para estas distribuciones son:

$$H(A\theta_{Sh}^*) = H(\omega_{Sh}^*) = 0.947$$

$$H(A\theta^*) = H(\omega^*) = 0.943.$$

Aunque la diferencia entre θ_{Sh}^* y θ^* puede parecer significativa, la diferencia entre los valores de la entropía que producen $A\theta_{Sh}^*$ y $A\theta^*$ es insignificante.

3.2.2. Análisis del error cometido

En algunas situaciones se puede considerar como solución del programa [I] la solución aproximada obtenida mediante el programa [II] independientemente de la entropía fijada ya que entonces el error cometido es despreciable. Tal es el caso de determinadas variedades lineales (que analizaremos posteriormente), o de aquellas medidas de entropía de la familia $\mathcal{H}$ cuyos conjuntos de nivel mantienen una cierta "esfericidad", es decir, uniformidad de las distancias entre los puntos que forman dichos conjuntos de nivel y $\omega^{(0)}$. También, si ω^* está próximo a $\omega^{(0)}$ el error será pequeño. Analizamos a continuación esta situación:

1. Fijado $r \geq k$.
2. Fijada la variedad lineal $\omega = A\theta$
3. Sea $\omega^* = A\theta^*$, es decir, el punto de la variedad lineal tal que

$$||\omega^* - \omega^{(0)}|| = \min_{\omega = A\theta} ||\omega - \omega^{(0)}||$$

4. Sea $\omega_H^* = A\theta_H^*$, es decir, el punto de la variedad lineal tal que, para la entropía $H \in \mathcal{H}$

$$\max_{\omega = A\theta} H(\omega) = H(\omega_H^*)$$

5. Supongamos que ω^* es un punto próximo a $\omega^{(0)}$, es decir $||\omega^* - \omega^{(0)}|| \leq \delta$

Vamos a estudiar el error cometido, en unidades de entropía, al elegir el punto ω^* en lugar de ω_H^*. Para ello vamos a acotar el valor de $H(\omega_H^*) - H(\omega^*)$. Teniendo en cuenta que esta diferencia depende de la entropía elegida H, parece más razonable estudiar el error relativo:

$$\triangle H^* = \frac{H(\omega_H^*) - H(\omega^*)}{H(\omega^{(0)})}$$

es decir, la pérdida relativa de entropía, en relación a la entropía máxima $H(\omega^{(0)})$.

Las situaciones que pueden presentarse, según las diferentes entropías son:

a) ω_H^* está alejado de ω^*

Esta situación sólo puede ocurrir si el crecimiento de H en la dirección del vector $\overline{\omega^*\omega_H^*}$ es muy lento en relación al crecimiento en la dirección del vector $\overline{\omega^*\omega^{(0)}}$.

Por la continuidad de la entropía H, existe un punto ω_H^A pertenenciente al segmento $[\omega^*, \omega^{(0)}]$ tal que $H(\omega_H^A) = H(\omega_H^*)$ es decir, el conjunto de nivel al que pertenece ω_H^* pasa por dicho punto y evidentemente se verifica que

$$||\omega^* - \omega_H^A|| < ||\omega^* - \omega^{(0)}|| \leq \delta$$

es decir, ω^* y ω_H^A son puntos próximos.

Fijado $\alpha > 0$, sea $\epsilon_H = \alpha H(\omega^{(0)}) > 0$, entonces, por la continuidad de H, existe un $\delta_H > 0$ tal que si

$$||\omega^* - \omega|| < \delta_H \Rightarrow \frac{|H(\omega) - H(\omega^*)|}{H(\omega^{(0)})} < \alpha$$

por tanto

$$\frac{H(\omega_H^A) - H(\omega^*)}{H(\omega^{(0)})} = \triangle H^* < \alpha \qquad \text{(por ejemplo } \alpha = 0.1)$$

para todo H, $\delta < \delta_H$.

b) ω_H^* está próximo a ω^*

$$||\omega^* - \omega_H^*|| \leq \delta$$

por la continuidad de H se tiene que

$$\triangle H^* = \frac{H(\omega_H^*) - H(\omega^*)}{H(\omega^{(0)})} < \alpha$$

para todo H tal que $\delta < \delta_H$.

Por otra parte si H es diferenciable, dada la proximidad entre ω_H^* y ω^* se pueden utilizar las aproximaciones del incremento de H que se deducen de la diferenciabilidad de H.

$$H(\omega_H^*) - H(\omega^*) \approx ||\nabla H(\omega^*)||\, ||\omega_H^* - \omega^*|| cos\alpha$$

con α el ángulo que forman los vectores $\nabla H(\omega^*)$ y $\overline{\omega^*\omega_H^*}$.

Si H es cóncava en $\triangle_k$

$$H(\omega_H^*) - H(\omega^*) \leq ||\nabla H(\omega^*)||\,||\omega_H^* - \omega^*||cos\alpha$$

luego

$$H(\omega_H^*) - H(\omega^*) \leq ||\nabla H(\omega^*)||\delta$$

de nuevo, la medida H elegida influye determinantemente en $\triangle H$.

Por último, tal como se comentaba al principio de esta sección, existen unas determinadas variedades lineales cuyas propiedades merece la pena comentar.

1. Sea $r > k$
2. Sea $\omega = A\theta$, la variedad lineal tal que $\omega_{i_1} = c_1, \omega_{i_2} = c_2, \ldots, \omega_{i_h} = c_h$ con $c_i > 0,\ i = 1, \ldots, h;\ \sum_{i=1}^{h} c_i < 1$ y $r - h \geq k$ (esta última desigualdad permite que la matriz A pueda ser de rango completo)

entonces el punto $\omega^{(1)}$ de componentes

$$\begin{aligned} \omega_{i_l}^{(1)} &= \frac{1 - \sum_{i=1}^{h} c_i}{r - h} \qquad l = h+1, \ldots, r \\ \omega_{i_l}^{(1)} &= c_l \qquad l = 1, \ldots, h \end{aligned}$$

es el punto más próximo a $\omega^{(0)}$ entre los que verifican la condición 2 anterior.
Si $\omega^{(1)} \in A\theta \Rightarrow \omega^{(1)} = \omega^* = \omega_H^*$ para cualquier entropía, por ser $\omega^{(1)}$ la "distribución uniforme" dentro de la variedad lineal.

De aquí, se deduce que para variedades lineales tales que uno o varios ω_i verifiquen:

$$c_i \leq \omega_i < c_i + \epsilon \qquad (\epsilon \text{ pequeño})$$

las soluciones ω^* y ω_H^* coincidirán prácticamente para toda entropía.

3.2.3. Ejemplos de acotación del error

Una cota para $\triangle H^*$, sencilla de calcular para cualquier entropía H, una vez fijada la variedad lineal $\omega = A\theta$, es la siguiente

$$\triangle H^* < \frac{H(\omega^{(0)}) - H(\omega^*)}{H(\omega^{(0)})} = \triangle H^{(0)}$$

Las tablas 1, 2 y 3 muestran para diferentes medidas de entropía y diferentes distancias $\delta = ||\omega^* - \omega^{(0)}||$, con $r = 3$, los valores máximos de $\triangle H^{(0)}$ para cualquier variedad lineal.

Interpretación: Fijado δ y el parámetro de la entropía, $\triangle H^{(0)}$ es menor o igual que la cantidad que aparece en la casilla correspondiente.

Tabla 1. Entropía de Rényi de parámetro t.

$max \triangle H^{(0)}$	$t = 1/2$	$t = 2$	$t = 3$	$t = 4$
$\delta = 0.1$	0.0065	0.02	0.041	0.054
$\delta = 0.2$	0.0347	0.102	0.147	0.189
$\delta = 0.3$	0.0910	0.213	0.295	0.345
$\delta = 0.4$	0.2690	0.353	0.447	0.501
$\delta = 0.5$	0.3560	0.509	0.595	0.639

Tabla 2. Entropía de Havrda y Charvát de parámetro s.

$max \triangle H^{(0)}$	$s = 1/2$	$s = 2$	$s = 3$	$s = 4$
$\delta = 0.1$	0.0086	0.013	0.013	0.0065
$\delta = 0.2$	0.0434	0.061	0.047	0.0326
$\delta = 0.3$	0.1173	0.135	0.101	0.0826
$\delta = 0.4$	0.3304	0.241	0.208	0.1611
$\delta = 0.5$	0.4261	0.375	0.339	0.273

Tabla 3. Entropía de Shannon.

δ	$max\triangle H^{(0)}$
$\delta = 0.1$	0.015
$\delta = 0.2$	0.062
$\delta = 0.3$	0.153
$\delta = 0.4$	0.331
$\delta = 0.5$	0.432

Observaciones

1. Diremos que una variedad lineal está alejada cuando ω^* no pertenezca a la bola cerrada de centro $\omega^{(0)}$ y radio $\sqrt{\frac{r-2}{2r}}$ siendo $\sqrt{\frac{r-2}{2r}}$ la distancia entre el punto $\omega^{(0)} = (1/r, \ldots, 1/r)$ y el punto medio del segmento que une cualquier par de vértices del conjunto Ω_r, por tanto, los resultados de las dos últimas filas corresponden a pérdidas relativas para variedades lineales alejadas ya que

$$\sqrt{\frac{r-2}{2r}} = \sqrt{\frac{1}{6}} = 0{,}4082.$$

2. Las cotas presentadas en las tablas son sólo eso. No quiere decir que $\triangle H^*$ tenga ese orden de magnitud. Si se fija la variedad lineal se pueden obtener los valores exactos de $\triangle H^*$ que serán inferiores a los de las tablas.

3. Analizadas las tablas se observa que en el caso de la entropía de Rényi (Tabla 1) es conveniente disminuir el valor del parámetro t para reducir el error $\triangle H^*$. En el caso de la entropía de Havrda y Charvát (Tabla 2) es conveniente aumentar el valor del parámetro s, $(s > 1)$ para reducir dicho error. Aumentar o disminuir, exageradamente, los valores de los parámetros ocasiona generalmente una pérdida de poder discriminante de las entropías frente a distribuciones "próximas" $\omega^{(1)}$, $\omega^{(2)}$.

Por otra parte, fijado $\alpha = 0{,}1$ la desigualdad

$$\frac{H(\omega^{(0)}) - H(\omega^*)}{H(\omega^{(0)})} < 0{,}1$$

se verifica para todo ω^* tal que:

1. H entropía de Shannon

   $$||\omega^* - \omega^{(0)}|| \leq 0{,}24$$

2. H entropía de Rényi $(1 < t \leq 5)$

   $$||\omega^* - \omega^{(0)}|| \leq 0{,}17$$

3. H entropía de Havrda y Chárvat $(2 < s \leq 6)$

   $$||\omega^* - \omega^{(0)}|| \leq 0{,}3$$

Conclusiones:

Fijado el problema a resolver, como el definido por el programa [I], se pueden estalecer las siguientes consideraciones:

1. Fijada la variedad lineal, la distribución de máxima entropía $\omega_H^* = A\theta_H^*$ depende de la medida de entropía escogida, en contra de la idea intuitiva de que fijada la variedad lineal, la máxima entropía es una cualidad de una determinada distribución y no depende del instrumento de medida elegido.

2. Para puntos ω^* próximos a $\omega^{(0)}$, la solución del programa [II], ω^* resulta una buena aproximación de la solución del programa [I], independientemente de la entropía elegida.

3. Se deben preferir las entropías cuyos conjuntos de nivel mantengan una cierta "esfericidad", es decir, uniformidad de las distancias entre los puntos que forman los conjuntos de nivel y $\omega^{(0)}$ pues en tal caso, se asegura la proximidad entre la solución del programa [I] y la solución del programa [II] independientemente de cual sea la variedad lineal fijada.

4. Para las entropías: Rényi de orden 2 y Havrda y Charvát de grado 2, la solución del programa [I] coincide con la del programa [II].

3.3. Formulación del programa

1. El problema de programación matemática que debemos resolver es el siguiente:

$$\left.\begin{array}{l} \text{mín } ||\omega^{(0)} - A\theta|| \\ s.a. \\ \theta_j \geq 0, \quad j = 1, \ldots, k \\ \displaystyle\sum_{j=1}^{k} \theta_j = 1 \end{array}\right\}$$

o de forma equivalente

$$\left.\begin{array}{l} \text{mín } ||\omega^{(0)} - A\theta||^2 \\ s.a. \\ \theta_j \geq 0, \quad j = 1, \ldots, k \\ \displaystyle\sum_{j=1}^{k} \theta_j = 1 \end{array}\right\} \qquad \text{[II]}$$

2. Características de la función objetivo y de la región factible.

a) Función objetivo

$$G(\theta) = ||\,\omega^{(0)} - A\,\theta\,||^2 = \sum_{i=1}^{r} \left(\frac{1}{r} - \sum_{j=1}^{k} a_{ij}\,\theta_j \right)^2$$

es una función continua y diferenciable en $\mathbb{R}^k$ como función de θ.

La matriz Hessiana

$$HG(\theta) = 2(h_{ij})_{i,j=1,\ldots,k}$$

es una matriz simétrica con

$$h_{ii} = \sum_{l=1}^{r} a_{li}^2 = a_{1i}^2 + a_{2i}^2 + \cdots + a_{ri}^2 \qquad i = 1, \ldots, k$$

$$h_{ij} = h_{ji} = \sum_{l=1}^{r} a_{li} a_{lj} \qquad i \neq j = 1, \ldots, k$$

La matriz $HG(\theta)$ puede escribirse:

$$HG(\theta) = 2\,A'\,A$$

Al ser $A'\,A$ una matriz semidefinida positiva, si A es de rango completo $A'\,A$ es definida positiva, lo que implica que $HG(\theta)$ será semidefinida positiva o definida positiva dependiendo del rango de A (Proposición A.2.2). Por tanto $G(\theta)$ es una función convexa en ambas situaciones, siendo estrictamente convexa cuando la matriz A sea de rango completo (Proposición 2.1.2).

b) El conjunto de soluciones factibles

$$\triangle_k = \{\, \theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k : \ \theta_j \geq 0,\ j = 1, \ldots, k,\ \sum_{j=1}^{k} \theta_j = 1 \,\}$$

es cerrado, acotado y convexo.

Demostración.

Es cerrado, pues su complementario es abierto, y acotado, pues existen bolas de radio finito que lo contienen (ver definición B.1.1). Demostraremos que es convexo utilizando la definición de conjunto convexo; es decir, tenemos que demostrar que para cada par de puntos $\theta^{(1)}\,,\ \theta^{(2)} \in \triangle_k$ y para todo $\lambda \in [0,1]$ se verifica que

$$\theta^{(3)} = \lambda\theta^{(1)} + (1-\lambda)\theta^{(2)} \in \triangle_k.$$

Sea $\theta^{(3)} = (\theta_1^{(3)}, \ldots, \theta_k^{(3)})$; $\theta^{(3)} = (\lambda\theta_1^{(1)} + (1-\lambda)\theta_1^{(2)}\,, \ldots,\ \lambda\theta_k^{(1)} + (1-\lambda)\theta_k^{(2)})$, luego $\theta_j^{(3)} = \lambda\theta_j^{(1)} + (1-\lambda)\theta_j^{(2)}$, $j = 1, \ldots, k$ de donde se desprende que $\theta_j^{(3)} \geq 0$ $j = 1, \ldots, k$ y como

$$\sum_{j=1}^{k} \theta_j^{(3)} = \lambda \sum_{j=1}^{k} \theta_j^{(1)} + (1-\lambda) \sum_{j=1}^{k} \theta_j^{(2)} = \lambda + (1-\lambda) = 1$$

queda demostrado que $\theta^{(3)} \in \triangle_k$ y por tanto $\triangle_k$ es convexo.

3. Consecuencias.

Como la función objetivo es convexa y el conjunto $\triangle_k$ es convexo, el problema de minimización [II] es convexo para mínimo y el teorema fundamental de la programación convexa (Teorema 2.2.3) garantiza que si θ^* es un mínimo local, entonces es un mínimo global, siendo convexo el conjunto de todos los mínimos del programa.

Puesto que $\triangle_k$ es un conjunto cerrado y acotado y $G(\theta)$ es continua en $\triangle_k$, el teorema de Weierstrass (Teorema 2.2.1) garantiza que $G(\theta)$ alcanza un valor mínimo en $\triangle_k$ luego el conjunto de soluciones del programa [II] es no vacío.

Las condiciones de Kuhn-Tucker caracterizan las soluciones globales en programas convexos, tanto de minimización como de maximización.

Todo θ^* que verifique las condiciones de Kuhn-Tucker será un mínimo global.

3.4. Resolución del programa

Como se trata de un programa convexo de minimización, las condiciones de Kuhn-Tucker son condiciones necesarias y suficientes para la existencia de óptimo global θ^* (no necesariamente único) que será único cuando la función objetivo sea estrictamente convexa.

Estas condiciones en nuestro caso son las siguientes:

$$(1) \quad \frac{\partial L}{\partial \theta_j}(\theta^*) = 0 \qquad j = 1, \ldots, k$$

$$(2) \quad \lambda_j \, \theta_j^* = 0 \qquad j = 1, \ldots, k$$

$$(3) \quad \theta_j^* \geq 0 \qquad j = 1, \ldots, k$$

$$(4) \quad \lambda_j \leq 0 \qquad j = 1, \ldots, k$$

$$(5) \quad \sum_{j=1}^{k} \theta_j^* = 1$$

siendo L la función Lagragiana:

$$L(\theta_1,\ldots,\theta_k,\ \lambda_1,\ldots,\lambda_k,\ \mu\,) = G(\theta) + \sum_{j=1}^{k}\lambda_j\theta_j + \mu\left(\sum_{j=1}^{k}\theta_j - 1\right)$$

Para resolver analíticamente el programa, hay que encontrar las soluciones del siguiente conjunto de ecuaciones formado por las condiciones de Kuhn-Tucker.

$$(1)\quad -2\left(\sum_{i=1}^{r}a_{ij}\left(\frac{1}{r}-\sum_{l=1}^{k}a_{il}\theta_l\right)\right)+\lambda_j+\mu=0 \qquad j=1,\ldots,k$$

$$(2)\quad \lambda_j\,\theta_j = 0 \qquad j=1,\ldots,k$$

$$(3)\quad \theta_j \geq 0 \qquad j=1,\ldots,k$$

$$(4)\quad \lambda_j \leq 0 \qquad j=1,\ldots,k$$

$$(5)\quad \sum_{j=1}^{k}\theta_j = 1$$

es decir, hay un total de $4k+1$ condiciones. Obsérvese que μ puede tomar cualquier valor en $\mathbb{R}$ por ser el mutiplicador correspondiente a una restricción de igualdad.

Las hipótesis que podemos hacer sobre los valores que toman los $\lambda_j,\quad j=1,\ldots,k$ atendiendo a las restricciones que se saturan (comenzando por el caso en el que no se satura ninguna restricción) se resumen en los siguientes 2^k-1 casos:

1. $\lambda_1=\lambda_2=\ldots=\lambda_k=0$

2. $\lambda_{i_1}=\lambda_{i_2}=\ldots=\lambda_{i_{k-1}}=0;\quad i_1<i_2<\ldots<i_{k-1}\in\{1,\ldots,k\};\quad \begin{pmatrix} k \\ k-1 \end{pmatrix}$ *casos*

3. $\lambda_{i_1}=\lambda_{i_2}=\ldots=\lambda_{i_{k-2}}=0;\quad i_1<i_2<\ldots<i_{k-2}\in\{1,\ldots,k\};\quad \begin{pmatrix} k \\ k-2 \end{pmatrix}$ *casos*

$\vdots$ $\qquad\qquad$ $\vdots$

$k-1$. $\lambda_{i_1}=\lambda_{i_2}=0;\qquad i_1<i_2\in\{1,\ldots,k\};\qquad \begin{pmatrix} k \\ 2 \end{pmatrix}$ *casos*

k. $\lambda_i=0;\qquad i\in\{1,\ldots,k\};\qquad \begin{pmatrix} k \\ 1 \end{pmatrix}$ *casos.*

$$\left(\begin{array}{c} k \\ 1 \end{array}\right) + \left(\begin{array}{c} k \\ 2 \end{array}\right) + \cdots + \left(\begin{array}{c} k \\ k-2 \end{array}\right) + \left(\begin{array}{c} k \\ k-1 \end{array}\right) + \left(\begin{array}{c} k \\ k \end{array}\right) = 2^k - 1.$$

En el caso $k+1$, se saturan todas las restricciones ($\lambda_1 \leq 0,\ \lambda_2 \leq 0, \cdots, \lambda_k \leq 0$) por tanto ha de verificarse el sistema:

$$\left.\begin{array}{l} \theta_1 = \theta_2 = \cdots = \theta_k = 0 \\ \theta_1 + \theta_2 + \cdots + \theta_k = 1 \end{array}\right\}$$

que carece de solución.

Dada la forma de las restricciones del programa, es fácil demostrar que se verifican las condiciones de regularidad (citadas en el capítulo anterior) en cada uno de los casos.

Nota.- En adelante se utilizará la notación matricial presentada en el apéndice A.

3.4.1. $rg(A) = k$, A matriz de rango completo

Para encontrar la solución del programa que va a ser única, pues la función objetivo es estrictamente convexa, y está grarantizada su existencia por el T. de Weierstrass, analizamos cada uno de los casos citados anteriormente, teniendo en cuenta que cuando encontremos un θ^* en alguno de estos casos que verifique las condiciones de Kuhn-Tucker no será necesario seguir.

Caso 1.

Si $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$ resulta el siguiente sistema de ecuaciones lineales, formado por las condiciones (1) y (5).

$$\left.\begin{array}{rcl}
-2\left[\left(\frac{a_{11}}{r} - a_{11}\sum_{j=1}^{k} a_{1j}\theta_j\right) + \left(\frac{a_{21}}{r} - a_{21}\sum_{j=1}^{k} a_{2j}\theta_j\right) + \cdots + \left(\frac{a_{r1}}{r} - a_{r1}\sum_{j=1}^{k} a_{rj}\theta_j\right)\right] + \mu & = & 0 \\
-2\left[\left(\frac{a_{12}}{r} - a_{12}\sum_{j=1}^{k} a_{1j}\theta_j\right) + \left(\frac{a_{22}}{r} - a_{22}\sum_{j=1}^{k} a_{2j}\theta_j\right) + \cdots + \left(\frac{a_{r2}}{r} - a_{r2}\sum_{j=1}^{k} a_{rj}\theta_j\right)\right] + \mu & = & 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & & \\
-2\left[\left(\frac{a_{1k}}{r} - a_{1k}\sum_{j=1}^{k} a_{1j}\theta_j\right) + \left(\frac{a_{2k}}{r} - a_{2k}\sum_{j=1}^{k} a_{2j}\theta_j\right) + \cdots + \left(\frac{a_{rk}}{r} - a_{rk}\sum_{j=1}^{k} a_{rj}\theta_j\right)\right] + \mu & = & 0 \\
\theta_1 + \theta_2 + \cdots + \theta_k & = & 1
\end{array}\right\}$$

y considerando que $\sum_{i=1}^{r} a_{ij} = 1, \;\; j = 1, \ldots, k$ llegamos a este otro sistema equivalente

$$\left.\begin{array}{rcl}
2\left(\sum_{i=1}^{r} a_{i1}^2\right)\theta_1 \;+\; 2\left(\sum_{i=1}^{r} a_{i1}a_{i2}\right)\theta_2 + \cdots + 2\left(\sum_{i=1}^{r} a_{i1}a_{ik}\right)\theta_k + \mu & = & \frac{2}{r} \\
2\left(\sum_{i=1}^{r} a_{i2}a_{i1}\right)\theta_1 + 2\left(\sum_{i=1}^{r} a_{i2}^2\right)\theta_2 + \cdots + 2\left(\sum_{i=1}^{r} a_{i2}a_{ik}\right)\theta_k + \mu & = & \frac{2}{r} \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots & & \\
2\left(\sum_{i=1}^{r} a_{ik}a_{i1}\right)\theta_1 + 2\left(\sum_{i=1}^{r} a_{ik}a_{i2}\right)\theta_2 + \cdots + 2\left(\sum_{i=1}^{r} a_{ik}^2\right)\theta_k + \mu & = & \frac{2}{r} \\
\theta_1 \;+\; \theta_2 \;+\; \cdots \;+\; \theta_k & = & 1
\end{array}\right\} \quad [1]$$

La matriz formada por los coeficientes de los θ_j, $j = 1, \cdots, k$ correspondientes a las k primeras ecuaciones coincide con la matriz hessiana $HG(\theta)$, es decir $2(h_{ij})_{i,j=1,\ldots,k}$ o $2A'A$.

Si llamamos a esta matriz H_k las matrices de coeficientes y ampliada del sistema [1] son:

$$M = \begin{pmatrix} H_k & 1_{k\times 1} \\ 1'_{k\times 1} & O \end{pmatrix} \quad ; \qquad M^* = \begin{pmatrix} H_k & 1_{k\times 1} & N_{k\times 1} \\ 1'_{k\times 1} & O & 1 \end{pmatrix}$$

siendo $1'_{k\times 1} = (\overbrace{1\ldots 1}^{k\,veces})$, $O = (0)$, $N'_{k\times 1} = (\overbrace{2/r\ldots 2/r}^{k\,veces})$ y $1 = (1)$

por tanto, M es una matriz cuadrada de orden $(k+1)\times(k+1)$ y M^* es una matriz de orden $(k+1)\times(k+2)$.

Estudio de la compatibilidad del sistema [1]. Resolución

Se trata de un sistema de ecuaciones lineales, por el teorema de Rouche-Frobenius el sistema es compatible si y sólo si $rg(M) = rg(M^*)$.

a) Estudio del rango de M

El rango de M viene determinado por el rango de A cuando A es de rango completo como se refleja en las siguientes proposiciones:

Proposición 3.4.1.

Dada la matriz

$$M = \begin{pmatrix} H_n & 1_{n\times 1} \\ 1'_{n\times 1} & O \end{pmatrix}$$

donde H_n es simétrica con elementos reales, $1'_{n\times 1} = (\overbrace{1\ldots 1}^{n\ veces})$, $O = (0)$.

Si H_n es definida positiva, entonces M es no singular.

Demostración.

Tenemos que demostrar que si H_n es definida positiva entonces $|M| \neq 0$. Por ser H_n definida positiva la forma cuadrática $q(\bar{x}) = \bar{x}' H_n \bar{x}$ con matriz asociada H_n es definida positiva, y por definición $q(\bar{x}) > 0$, para todo $\bar{x} \in \mathbb{R}^n$, $\bar{x} \neq 0$ por tanto $q(\bar{x})$ es definida positiva en cualquier subconjunto de $\mathbb{R}^n$; en particular, en el conjunto

$$S = \{\bar{x} \in \mathbb{R}^n / \, x_1 + x_2 + \cdots + x_n = 0\} = \{\bar{x} \in \mathbb{R}^n / B\bar{x} = 0\}$$

con $B = (1\ldots 1)$, luego la forma cuadrática restringida $q(\bar{x}) = \bar{x}' H_n \bar{x}$ sujeta a $B\bar{x} = 0$ es definida positiva.

Por el lema A.2.1 (Apén. A) (obsérvese que aquí $m = 1$ y $rg(B) = 1$) existe una forma cuadrática $q^*(\bar{y}) = \bar{y}\,' E \bar{y}$ con $\bar{y} \in \mathbb{R}^{n-1}$ con matriz asociada E, tal que q^* es definida positiva, lo que nos permite afirmar que $|E| \neq 0$.

Por el Lema A.2.2, en nuestro caso para $i = n - 1$, se obtiene que $|M| = -1 \cdot 1 \cdot |E|$ y al ser $|E| \neq 0 \Rightarrow |M| \neq 0$, como queríamos demostrar.

Proposición 3.4.2.

Dada la matriz

$$M = \begin{pmatrix} H_n & 1_{n\times 1} \\ 1'_{n\times 1} & O \end{pmatrix}$$

donde H_n es simétrica con elementos reales, $1'_{n\times 1} = (\overbrace{1\ldots 1}^{n\ veces})$, $O = (0)$.

Si $H_n = 2B'B$ y B es de rango completo, entonces M es no singular.

Demostración

Tenemos que demostrar que si B es de rango completo $|M| \neq 0$. Por ser B de rango completo $H_n = 2B'B$ es definida positiva (Proposición A.2.2) luego aplicando la proposición anterior queda demostrado que M es no singular.

En nuestro caso $rg(M) = k + 1$

b) Estudio del rango de M^*

El rango de la matriz M^* es $k + 1$ pues existe un menor de orden $k + 1$ formado por sus primeras $k + 1$ columnas distinto de 0.

c) Resolución

Como $rg(M) = rg(M^*) = k + 1 =$ número de incógnitas, el sistema es compatible determinado, por tanto la solución es única pudiéndose obtener ésta por las conocidas fórmulas de Cramer:

$$\theta_j^* = \frac{|M_{\theta_j}|}{|M|} \qquad j = 1, \ldots, k$$

$$\mu^* = \frac{|M_\mu|}{|M|}$$

siendo:

M_{θ_j} la matriz que se obtiene de M reemplazando la columna j-ésima $j = 1, \ldots, k$ por la columna de términos independientes

M_μ la matriz que se obtiene de M reemplazando la columna $(k + 1)$-ésima por la columna de términos independientes.

Necesitamos finalmente verificar que estos θ_j^*, $j = 1, \ldots, k$ cumplen las condiciones de Kuhn-Tucker (μ puede tomar cualquier valor en $\mathbb{R}$) que en nuestro caso se reduce a la condición (3) es decir $\theta_j \geq 0$, $j = 1, \ldots, k$.

Caso 2. $\lambda_{i_1} = \lambda_{i_2} = \cdots = \lambda_{i_{k-1}} = 0; \qquad i_1 < i_2 < \ldots < i_{k-1} \in \{1, \ldots, k\}$

Supongamos, sin pérdida de generalidad, que $\lambda_{i_h} = \lambda_h, \quad h = 1, \ldots, k-1$. El sistema de ecuaciones lineales que resulta teniendo en cuenta las condiciones (1), (2) y (5) de Kuhn-Tucker, es el siguiente:

$$\left.\begin{array}{rcl}
-2\left[\left(\dfrac{a_{11}}{r} - a_{11}\displaystyle\sum_{j=1}^{k-1} a_{1j}\theta_j\right) + \left(\dfrac{a_{21}}{r} - a_{21}\displaystyle\sum_{j=1}^{k-1} a_{2j}\theta_j\right) + \cdots + \left(\dfrac{a_{r1}}{r} - a_{r1}\displaystyle\sum_{j=1}^{k-1} a_{rj}\theta_j\right)\right] + \mu & = & 0 \\
-2\left[\left(\dfrac{a_{12}}{r} - a_{12}\displaystyle\sum_{j=1}^{k-1} a_{1j}\theta_j\right) + \left(\dfrac{a_{22}}{r} - a_{22}\displaystyle\sum_{j=1}^{k-1} a_{2j}\theta_j\right) + \cdots + \left(\dfrac{a_{r2}}{r} - a_{r2}\displaystyle\sum_{j=1}^{k-1} a_{rj}\theta_j\right)\right] + \mu & = & 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & & \\
-2\left[\left(\dfrac{a_{1k}}{r} - a_{1k}\displaystyle\sum_{j=1}^{k-1} a_{1j}\theta_j\right) + \left(\dfrac{a_{2k}}{r} - a_{2k}\displaystyle\sum_{j=1}^{k-1} a_{2j}\theta_j\right) + \cdots + \left(\dfrac{a_{rk}}{r} - a_{rk}\displaystyle\sum_{j=1}^{k-1} a_{rj}\theta_j\right)\right] + \lambda_k + \mu & = & 0 \\
\theta_1 + \theta_2 + \cdots + \theta_{k-1} & = & 1
\end{array}\right\}$$

y considerando que $\displaystyle\sum_{i=1}^{r} a_{ij} = 1, \quad j = 1, \ldots, k,$ llegamos a este otro sistema equivalente

$$\left.\begin{array}{rcl}
2\left(\displaystyle\sum_{i=1}^{r} a_{i1}^2\right)\theta_1 + 2\left(\displaystyle\sum_{i=1}^{r} a_{i1}a_{i2}\right)\theta_2 + \cdots + 2\left(\displaystyle\sum_{i=1}^{r} a_{i1}a_{ik-1}\right)\theta_{k-1} + \mu & = & \dfrac{2}{r} \\
2\left(\displaystyle\sum_{i=1}^{r} a_{i2}a_{i1}\right)\theta_1 + 2\left(\displaystyle\sum_{i=1}^{r} a_{i2}^2\right)\theta_2 + \cdots + 2\left(\displaystyle\sum_{i=1}^{r} a_{i2}a_{ik-1}\right)\theta_{k-1} + \mu & = & \dfrac{2}{r} \\
\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots & & \\
2\left(\displaystyle\sum_{i=1}^{r} a_{ik}a_{i1}\right)\theta_1 + 2\left(\displaystyle\sum_{i=1}^{r} a_{ik}a_{i2}\right)\theta_2 + \cdots + 2\left(\displaystyle\sum_{i=1}^{r} a_{ik}a_{ik-1}\right)\theta_{k-1} + \lambda_k + \mu & = & \dfrac{2}{r} \\
\theta_1 + \theta_2 + \cdots + \theta_{k-1} & = & 1
\end{array}\right\} \quad [2]$$

La matriz formada por los coeficientes de los $\theta_j, \ j = 1, \ldots, k-1,$ correspondientes a las k primeras ecuaciones, es una submatriz de la matriz hessiana H_k que se obtiene eliminando la columna k-ésima de la msima. Si llamamos a esta matriz $H_{k\times(k-1)}$ y la

particionamos de la siguiente forma:

$$H_{k\times(k-1)} = \begin{pmatrix} H_{(k-1)} \\ H_{1\times(k-1)} \end{pmatrix}$$

con $H_{(k-1)}$ matriz cuadrada de orden $(k-1)$ formada por las $k-1$ primeras filas de $H_{k\times(k-1)}$ (en definitiva, $H_{(k-1)}$ es la matriz que resulta de suprimir la última fila y columna de H_k) y $H_{1\times(k-1)}$ la matriz de orden $1\times(k-1)$ formada por la fila k-ésima de $H_{k\times(k-1)}$, entonces las matrices de coeficientes y ampliada del sistema [2] son, respectivamente

$$M = \begin{pmatrix} H_{(k-1)} & O_{(k-1)\times 1} & 1_{(k-1)\times 1} \\ H_{1\times(k-1)} & 1 & 1 \\ 1'_{(k-1)\times 1} & O & O \end{pmatrix},$$

$$M^* = \begin{pmatrix} H_{(k-1)} & O_{(k-1)\times 1} & 1_{(k-1)\times 1} & N_{(k-1)\times 1} \\ H_{1\times(k-1)} & 1 & 1 & N_1 \\ 1'_{(k-1)\times 1} & O & O & 1 \end{pmatrix}$$

siendo $O_{(k-1)\times 1}$ la matriz nula de orden $(k-1)\times 1$, $N'_{(k-1)\times 1} = (\overbrace{2/r\cdots 2/r}^{(k-1\ veces)})$,

$N_1 = (2/r)$; $1'_{(k-1)\times 1} = (\overbrace{1,\ldots,1}^{(k-1)})$, $1 = (1)$ y $O = (0)$.

M es por tanto, una matriz cuadrada de orden $(k+1)\times(k+1)$ y M^* es una matriz de orden $(k+1)\times(k+2)$.

Estudio de la compatibilidad del sistema [2]. Resolución

Se trata de un sistema de ecuaciones lineales, por el teorema de Rouche-Frobenius, el sistema es compatible si y sólo si $rg(M) = rg(M^*)$.

a) Estudio del rango de M

En este caso el rango de M es $k+1$.

Demostración. Tenemos que demostrar que $|M| \neq 0$. Desarrollando dicho determinante por la columna k-ésima, resulta que

$$|M| = (-1)^{2k} \begin{vmatrix} H_{(k-1)} & 1_{(k-1)\times 1} \\ 1'_{(k-1)\times 1} & 0 \end{vmatrix}$$

$H_{(k-1)}$ es definida positiva por serlo H_k (proposición A.2.1) luego por la proposición 3.4.1. M es no singular por lo que $rg(M) = k + 1$.

b) Estudio del rango de M^*

El rango de la matriz M^* es $k + 1$ pues existe un menor de orden $k + 1$ formado por sus primeras $k + 1$ columnas distinto de 0.

c) Resolución

Como $rg(M) = rg(M^*) = k + 1 =$ número de incógnitas, el sistema es compatible determinado, por tanto, la solución es única pudiéndose obtener ésta por las conocidas fórmulas de Cramer:

$$\theta_j^* = \frac{|M_{\theta_j}|}{|M|} \qquad j = 1, \ldots, k-1$$

$$\lambda_k^* = \frac{|M_{\lambda_k}|}{|M|}$$

$$\mu^* = \frac{|M_\mu|}{|M|}$$

siendo:

M_{θ_j} la matriz que se obtiene de M reemplazando la columna j-ésima, $j = 1, \ldots, k-1$ por la columna de términos independientes

M_{λ_k} la matriz que se obtiene de M reemplazando la columna k-ésima por la columna de términos independientes

M_μ la matriz que se obtiene de M reemplazando la columna $(k + 1)$-ésima por la columna de términos independientes.

Necesitamos finalmente verificar que se cumplen las condiciones de Kuhn-Tucker que en este caso son las condiciones (3) y (4), $\theta_j^* \geq 0, \quad j = 1, \ldots, k-1,$ y $\lambda_k^* \leq 0$ (μ puede tomar cualquier valor en $\mathbb{R}$).

Para cualquier otra reordenación de los $\lambda_{i_h} = 0, \ h = 1, \ldots, k-1,$ siguen siendo válidas las conclusiones sobre el estudio de la compatibilidad y resolución del sistema [2], obtenidas anteriormente ya que, aunque la matriz M no esté particionada de la misma forma en la que aparece anteriormente, se sigue verificando que $|M| \neq 0$ por las mismas razones expuestas, basta desarrollar dicho determinante por la columna adecuada para cada caso.

Caso 3. $\lambda_{i_1} = \lambda_{i_2} = \ldots = \lambda_{i_{k-2}} = 0; \qquad i_1 < i_2 < \ldots < i_{k-2} \in \{1, \ldots, k\}$

Supongamos, sin pérdida de generalidad, que $\lambda_{i_h} = \lambda_h,\ h = 1, \ldots, k-2$. El sistema de ecuaciones lineales que resulta teniendo en cuenta las condiciones (1), (2) y (5) de Kuhn-Tucker, es el siguiente:

$$\left.\begin{array}{rcl}
-2\left[\left(\frac{a_{11}}{r} - a_{11}\sum_{j=1}^{k-2} a_{1j}\theta_j\right) + \left(\frac{a_{21}}{r} - a_{21}\sum_{j=1}^{k-2} a_{2j}\theta_j\right) + \cdots + \left(\frac{a_{r1}}{r} - a_{r1}\sum_{j=1}^{k-2} a_{rj}\theta_j\right)\right] + \mu & = & 0 \\
-2\left[\left(\frac{a_{12}}{r} - a_{12}\sum_{j=1}^{k-2} a_{1j}\theta_j\right) + \left(\frac{a_{22}}{r} - a_{22}\sum_{j=1}^{k-2} a_{2j}\theta_j\right) + \cdots + \left(\frac{a_{r2}}{r} - a_{r2}\sum_{j=1}^{k-2} a_{rj}\theta_j\right)\right] + \mu & = & 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots & & \\
-2\left[\left(\frac{a_{1k-1}}{r} - a_{1k-1}\sum_{j=1}^{k-2} a_{1j}\theta_j\right) + \left(\frac{a_{2k-1}}{r} - a_{2k-1}\sum_{j=1}^{k-2} a_{2j}\theta_j\right) + \cdots + \left(\frac{a_{rk-1}}{r} - a_{rk-1}\sum_{j=1}^{k-2} a_{rj}\theta_j\right)\right] + \lambda_{k-1} + \mu & = & 0 \\
-2\left[\left(\frac{a_{1k}}{r} - a_{1k}\sum_{j=1}^{k-2} a_{1j}\theta_j\right) + \left(\frac{a_{2k}}{r} - a_{2k}\sum_{j=1}^{k-2} a_{2j}\theta_j\right) + \cdots + \left(\frac{a_{rk}}{r} - a_{rk}\sum_{j=1}^{k-2} a_{rj}\theta_j\right)\right] + \lambda_k + \mu & = & 0 \\
\theta_1 + \theta_2 + \cdots + \theta_{k-2} & = & 1
\end{array}\right\}$$

y considerando que $\sum_{i=1}^{r} a_{ij} = 1,\ \ j = 1, \ldots, k,$ llegamos a este otro sistema equivalente

$$\left.\begin{array}{rcl}
2\left(\sum_{i=1}^{r} a_{i1}^2\right)\theta_1 + 2\left(\sum_{i=1}^{r} a_{i1}a_{i2}\right)\theta_2 + \cdots + 2\left(\sum_{i=1}^{r} a_{i1}a_{ik-2}\right)\theta_{k-2} + \mu & = & \frac{2}{r} \\
2\left(\sum_{i=1}^{r} a_{i2}a_{i1}\right)\theta_1 + 2\left(\sum_{i=1}^{r} a_{i2}^2\right)\theta_2 + \cdots + 2\left(\sum_{i=1}^{r} a_{i2}a_{ik-2}\right)\theta_{k-2} + \mu & = & \frac{2}{r} \\
\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots & & \\
2\left(\sum_{i=1}^{r} a_{ik-1}a_{i1}\right)\theta_1 + 2\left(\sum_{i=1}^{r} a_{ik-1}a_{i2}\right)\theta_2 + \cdots + 2\left(\sum_{i=1}^{r} a_{ik-1}a_{ik-2}\right)\theta_{k-2} + \lambda_{k-1} + \mu & = & \frac{2}{r} \\
2\left(\sum_{i=1}^{r} a_{ik}a_{i1}\right)\theta_1 + 2\left(\sum_{i=1}^{r} a_{ik}a_{i2}\right)\theta_2 + \cdots + 2\left(\sum_{i=1}^{r} a_{ik}a_{ik-2}\right)\theta_{k-2} + \lambda_k + \mu & = & \frac{2}{r} \\
\theta_1 + \theta_2 + \cdots + \theta_{k-2} & = & 1
\end{array}\right\} \quad [3]$$

La matriz formada por los coeficientes de los $\theta_j,\ \ j = 1, \ldots, k-2,$ correspondiente a

las k primeras ecuaciones, es una submatriz de la matriz hessiana H_k que se obtiene eliminando las dos últimas columnas de la misma. Si llamamos a esta matriz $H_{k\times(k-2)}$ y la particionamos de la siguiente forma

$$H_{k\times(k-2)} = \begin{pmatrix} H_{(k-2)} \\ H_{2\times(k-2)} \end{pmatrix}$$

con $H_{(k-2)}$ matriz cuadrada de orden $(k-2)$ formada por las $k-2$ primeras filas de $H_{k\times(k-2)}$ y $H_{2\times(k-2)}$ la matriz de orden $2\times(k-2)$ formada por las dos últimas filas de $H_{k\times(k-2)}$, entonces las matrices de coeficientes y ampliada del sistema [3] son, respectivamente

$$M = \begin{pmatrix} H_{(k-2)} & O_{(k-2)\times 2} & 1_{(k-2)\times 1} \\ H_{2\times(k-2)} & I_2 & 1_{2\times 1} \\ 1'_{(k-2)\times 1} & O_{1\times 2} & O \end{pmatrix},$$

$$M^* = \begin{pmatrix} H_{(k-2)} & O_{(k-2)\times 2} & 1_{(k-2)\times 1} & N_{(k-2)\times 1} \\ H_{2\times(k-2)} & I_2 & 1_{2\times 1} & N_{2\times 1} \\ 1'_{(k-2)\times 1} & O_{1\times 2} & O & 1 \end{pmatrix}$$

siendo I_2 la matriz identidad de orden 2, $O_{j\times 2}$ la matriz nula de orden $j\times 2$, $N'_{j\times 1} = (\overbrace{2/r\cdots 2/r}^{j\ veces})$, $1'_{j\times 1} = (\overbrace{1\cdots 1}^{j\ veces})$, $1 = (1)$ y $O = (0)$.

M es por tanto, una matriz cuadrada de orden $(k+1)\times(k+1)$ y M^* es una matriz de orden $(k+1)\times(k+2)$.

Estudio de la compatibilidad del sistema [3]. Resolución.

a) Estudio del rango de M

En este caso el rango de M es $k+1$

Demostración. Tenemos que demostrar que $|M| \neq 0$. Desarrollando dicho determinante

por las columnas $k-1$ y k, resulta que

$$|M| = (-1)^{4(k-1)} \begin{vmatrix} H_{(k-2)} & 1_{(k-2)\times 1} \\ 1'_{(k-2)\times 1} & O \end{vmatrix}$$

$H_{(k-2)}$ es definida positiva por serlo H_k (proposición A.2.1), luego por la proposición 3.4.1. M es no singular por lo que $rg(M) = k+1$.

b) Estudio del rango de M^*

El rango de la matriz M^* es $k+1$ pues existe un menor de orden $k+1$ formado por sus primeras $k+1$ columnas distinto de 0.

c) Resolución

Como $rg(M) = rg(M^*) = k+1 =$ número de incógnitas, el sistema es compatible determinado, por tanto, la solución es única y se puede obtener por las conocidas fórmulas de Cramer

$$\theta_j^* = \frac{|M_{\theta_j}|}{|M|} \qquad j = 1, \ldots, k-2$$

$$\lambda_j^* = \frac{|M_{\lambda_j}|}{|M|} \qquad j = k-1,\ k$$

$$\mu^* = \frac{|M_\mu|}{|M|}$$

siendo:

M_{θ_j} la matriz que se obtiene de M reemplazando la columna j-ésima, $j = 1, \ldots, k-2$ por la columna de términos independientes

M_{λ_j} la matriz que se obtiene de M reemplazando la columna j-ésima, $j = k-1, k$ por la columna de términos independientes

M_μ la matriz que se obtiene de M reemplazando la columna $(k+1)$-ésima por la columna de términos independientes.

Necesitamos finalmente verificar que se cumplen las condiciones de Kuhn-Tucker que en

este caso son las condiciones (3) y (4), $\theta_j^* \geq 0, \;\; j = 1, \ldots, k-2,$ y $\lambda_j^* \leq 0, \;\; j = k-1, k$ (μ puede tomar cualquier valor en $\mathbb{R}$).

Para cualquier otra reordenación de los $\lambda_{i_h} = 0, \; h = 1, \ldots, k-2,$ siguen siendo válidas las conclusiones sobre el estudio de la compatibilidad y resolución del sistema [3], obtenidas anteriormente ya que, aunque la matriz M no esté particionada de la misma forma en la que aparece anteriormente, se sigue verificando que $|M| \neq 0$ por las mismas razones expuestas, basta desarrollar dicho determinante por las columnas adecuadas para cada caso.

***Casos* 4 al *k* −1.**

$\lambda_{i_1} = \lambda_{i_2} = \cdots = \lambda_{i_{k-l}} = 0, \quad i_1 < i_2 < \ldots < i_{k-l} \in \{1, \ldots, k\}, \quad l = 3, \ldots, k-2$

Supongamos, sin pérdida de generalidad, que $\lambda_{i_h} = \lambda_h, \ h = 1, \ldots, k-l$. El sistema de ecuaciones lineales que resulta teniendo en cuenta las condiciones (1), (2) y (5) de Kuhn-Tucker, es el siguiente:

$$\left.\begin{array}{rcl}
2\left(\sum_{i=1}^{r} a_{i1}^2\right)\theta_1 + \quad 2\left(\sum_{i=1}^{r} a_{i1}a_{i2}\right)\theta_2 + \cdots + \quad 2\left(\sum_{i=1}^{r} a_{i1}a_{ik-l}\right)\theta_{k-l} \qquad\qquad + \mu &=& \frac{2}{r} \\
2\left(\sum_{i=1}^{r} a_{i2}a_{i1}\right)\theta_1 + \quad 2\left(\sum_{i=1}^{r} a_{i2}^2\right)\theta_2 + \cdots + \quad 2\left(\sum_{i=1}^{r} a_{i2}a_{ik-l}\right)\theta_{k-l} \qquad\qquad + \mu &=& \frac{2}{r} \\
\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots & & \\
2\left(\sum_{i=1}^{r} a_{ik-l+1}a_{i1}\right)\theta_1 + 2\left(\sum_{i=1}^{r} a_{ik-l+1}a_{i2}\right)\theta_2 + \cdots + 2\left(\sum_{i=1}^{r} a_{ik-l+1}a_{ik-l}\right)\theta_{k-l} + \lambda_{k-l+1} \quad + \mu &=& \frac{2}{r} \\
\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots & & \\
2\left(\sum_{i=1}^{r} a_{ik-1}a_{i1}\right)\theta_1 + \quad 2\left(\sum_{i=1}^{r} a_{ik-1}a_{i2}\right)\theta_2 + \cdots + \quad 2\left(\sum_{i=1}^{r} a_{ik-1}a_{ik-l}\right)\theta_{k-l} + \qquad \lambda_{k-1} \quad + \mu &=& \frac{2}{r} \\
2\left(\sum_{i=1}^{r} a_{ik}a_{i1}\right)\theta_1 + \quad 2\left(\sum_{i=1}^{r} a_{ik}a_{i2}\right)\theta_2 + \cdots + \quad 2\left(\sum_{i=1}^{r} a_{ik}a_{ik-l}\right)\theta_{k-l} + \qquad \lambda_k + \mu &=& \frac{2}{r} \\
\theta_1 + \qquad\qquad \theta_2 \quad + \qquad\qquad \cdots \qquad\qquad + \quad \theta_{k-l} \qquad\qquad\qquad\qquad &=& 1
\end{array}\right\} [4]$$

La matriz formada por los coeficientes de los $\theta_j, \ j = 1, \ldots, k-l,$ correspondiente a las k primeras ecuaciones, es una submatriz de la matriz hessiana H_k que se obtiene eliminando las l últimas columnas de H. Si llamamos a esta matriz $H_{k\times(k-l)}$ y la particionamos de la siguiente forma:

$$H_{k\times(k-l)} = \begin{pmatrix} H_{(k-l)} \\ H_{l\times(k-l)} \end{pmatrix}$$

con $H_{(k-l)}$ matriz cuadrada de orden $(k-l)$ formada por las $k-l$ primeras filas de $H_{k\times(k-l)}$ y $H_{l\times(k-l)}$ la matriz de orden $l \times (k-l)$ formada por las l últimas filas de $H_{k\times(k-l)}$, entonces las matrices de coeficientes y ampliada del sistema anterior son, respectivamente

$$M = \begin{pmatrix} H_{(k-l)} & O_{(k-l)\times l} & 1_{(k-l)\times 1} \\ H_{l\times(k-l)} & I_l & 1_{l\times 1} \\ 1'_{(k-l)\times 1} & O_{1\times l} & O \end{pmatrix},$$

$$M^* = \begin{pmatrix} H_{(k-l)} & O_{(k-l)\times l} & 1_{(k-l)\times 1} & N_{(k-l)\times 1} \\ H_{l\times(k-l)} & I_l & 1_{l\times 1} & N_{l\times 1} \\ 1'_{(k-l)\times 1} & O_{1\times l} & O & 1 \end{pmatrix}$$

siendo I_l la matriz identidad de orden l, $O_{j\times l}$ la matriz nula de orden $j \times l$,
$N'_{j\times 1} = (\overbrace{2/r \cdots 2/r}^{j\ veces})$, $1'_{j\times 1} = (\overbrace{1, \ldots, 1}^{j\ veces})$, $1 = (1)$ y $O = (0)$.

M, es por tanto una matriz cuadrada de orden $(k+1)\times(k+1)$ y M^* es una matriz de orden $(k+1)\times(k+2)$.

Estudio de la compatibilidad del sistema [4]. Resolución

a) Estudio del rango de M

El rango de M es $k+1$ para $l = 3, \ldots, k-2$.

Demostración. Tenemos que demostrar que $|M| \neq 0$. para $l = 3, \ldots, k-2$. Desarrollando dicho determinante por las columnas $k-l+1, \ldots, k$, resulta que

$$|M| = (-1)^{2l(k-1)} \begin{vmatrix} H_{(k-l)} & 1_{(k-l)\times 1} \\ 1'_{(k-l)\times 1} & O \end{vmatrix}$$

$H_{(k-l)}$ es definida psositiva por serlo H_k luego, por la proposición 3.4.1. M es no singular por lo que $rg(M) = k+1$.

b) Estudio del rango de M^*

El rango de la matriz M^* es $k+1$ pues existe un menor de orden $k+1$ formado por sus primeras $k+1$ columnas distinto de 0.

c) Resolución

Como $rg(M) = rg(M^*) = k+1 =$ número de incógnitas, el sistema es compatible determinado, por tanto la solución es única pudiéndose obtener ésta por las conocidas fórmulas de Cramer

$$\theta_j^* = \frac{\left|M_{\theta_j}\right|}{|M|} \qquad j = 1, \ldots, k-l, \quad l = 3, \ldots, k-1$$

$$\lambda_j^* = \frac{\left|M_{\lambda_j}\right|}{|M|} \qquad j = k-l+1, \ldots, k, \quad l = 3, \ldots, k-1$$

$$\mu^* = \frac{|M_\mu|}{|M|}$$

siendo:

M_{θ_j} la matriz que se obtiene de M reemplazando la columna j-ésima, $j = 1, \ldots, k-l$ por la columna de términos independientes

M_{λ_j} la matriz que se obtiene de M reemplazando la columna j-ésima, $j = k-l+1, \ldots, k$ por la columna de términos independientes

M_μ la matriz que se obtiene de M reemplazando la columna $(k+1)$-ésima por la columna de términos independientes.

Necesitamos finalmente verificar que se cumplen las condiciones de Kuhn-Tucker que en este caso son las condiciones (3) y (4), $\theta_j^* \geq 0, \quad j = 1, \ldots, k-l$, y $\lambda_j^* \leq 0, \quad j = k-l+1, \ldots, k$ (μ puede tomar cualquier valor en $\mathbb{R}$).

Para cualquier otra reordenación de los λ_{i_h}, $h = 1, \ldots, k-l$, siguen siendo válidas las conclusiones sobre el estudio de la compatibilidad y resolución del sistema [4], obtenidas anteriormente ya que, aunque la matriz M no esté particionada de la misma forma en la que aparece anteriormente, se sigue verificando que $|M| \neq 0$ por las mismas razones expuestas, basta desarrollar dicho determinante por las columnas adecuadas para cada caso.

Caso *k*. $\lambda_j = 0, \qquad j \in \{1, \ldots, k\}$

El caso k es especial pues supone que solamente existe un $\theta_h \neq 0$ y que por la condición (5) de Kuhn-Tucker, va a ser $\theta_h = 1$ (distribución degenerada). Para averiguar que θ_j de entre $\theta_1, \ldots, \theta_k$ es el adecuado, tenemos dos opciones:

La primera se basa en que

$$\omega = A\,\theta$$

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_r \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rk} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{pmatrix}$$

es decir, $\omega' = (a_{1j}\theta_j, a_{2j}\theta_j, \ldots, a_{rj}\theta_j)' = (a_{1j}, a_{2j}, \ldots, a_{rj})' = a^{(j)}$, vector cuyas componentes son los elementos de la columna j-ésima de A, luego, bajo las condiciones señaladas, a la solución del programa [II] se llega a través de las columnas de la matriz A, precisamente de aquella que haga mínimo $||\omega^{(0)} - a^{(j)}||$, $j = 1, \ldots, k$, por tanto no es necesario resolver ningún sistema. Se calculan los valores de $||\omega^{(0)} - a^{(j)}||$ para $j = 1, \ldots, k$ y si el menor valor de los calculados corresponde a la columna h-ésima, $h \in \{1, \ldots, k\}$ se toma $\theta_h = 1$.

El segundo utiliza las condiciones de Kuhn-Tucker de la misma forma que en los casos anteriores, luego consiste en ir resolviendo para $j = 1, \ldots, k$ los sistemas de k ecuaciones con k incógnitas $(\lambda_1, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_k, \mu)$, expresados en forma vectorial como sigue:

$$2\bar{h}_{\bullet j} + \bar{\lambda}_j + \bar{\mu} = N_{k\times 1} \qquad [5]$$

siendo $2\bar{h}_{\bullet j}$ $j = 1, \ldots, k$ el vector cuyas componentes son los elementos de la columna j-ésima de la matriz $2A'A$; $\bar{\lambda}'_j = (\lambda_1, \ldots, \lambda_{j-1}, 0, \lambda_{j+1}, \ldots, \lambda_k)$, $j = 1, \ldots, k$, $\bar{\mu}' = (\overbrace{\mu, \ldots, \mu}^{k \text{ veces}})$ y $N'_{k\times 1} = (\overbrace{2/r, \ldots, 2/r}^{k \text{ veces}})$.

El sistema [5] que es compatible deteminado aporta los canditados a solución del programa [II], solución que se encuentra cuando se verifica la condición (4) de Kuhn-Tucker es decir $\bar{\lambda} \leq \bar{0}$.

3.4.2. Caso particular

Si la matriz A es cuadrada y no singular, existe A^{-1} y es posible buscar la solución $\theta^* \in \triangle_k$ del programa de una forma alternativa mucho más directa y que consiste en resolver el sistema compatible determinado $\omega^{(0)} = A\theta$ o expresado en forma matricial

$$\begin{pmatrix} 1/k \\ \vdots \\ 1/k \end{pmatrix} = A \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix} \qquad [6]$$

En definitiva lo que se busca es el $\theta^* \in \triangle_k$ que se transforma en la distribución uniforme, pues cualquier entropía alcanza su valor máximo absoluto con esta distribución y por tanto todo $\theta^* \in \triangle_k$ que verifique esta condición se convierte automáticamente en el punto óptimo que buscamos.

La solución del sistema [6] viene dada por:

$$\theta^* = A^{-1} \begin{pmatrix} 1/k \\ \vdots \\ 1/k \end{pmatrix}$$

sólo queda comprobar que $\theta^* \in \triangle_k$, pues, aunque A es la matriz de una transformación lineal de $\triangle_k$ en $\triangle_k$, no está garantizado que la imagen inversa de un elemento de $\Omega_k = \triangle_k$ pertenezca a $\triangle_k$; basta tener en cuenta para argumentar esta afirmación que en la construcción de esta imagen inversa interviene una matriz inversa en la que pueden aparecer elementos negativos. ¿Cuándo podemos afirmar que la imagen inversa de la distribución uniforme pertenece a $\triangle_k$? La respuesta se encuentra en la siguiente proposición

Proposición 3.4.3.

Sea A una matriz definida como en la sección 3.1, cuadrada y no singular y sea $A^{-1} = (\widehat{a}_{ij})_{i,j=1,\ldots,k}$ su matriz inversa, entonces $A^{-1}\omega^{(0)} \in \triangle_k$ si y sólo si para cada $i = 1, \ldots, k$ se verifica que $0 \leq \sum_{j=1}^{k} \widehat{a}_{ij} \leq k$.

Demostración.

Supongamos que $A^{-1}\omega^{(0)} \in \triangle_k$, entonces existe un $\theta = (\theta_1, \ldots, \theta_k) \in \triangle_k$ tal que

$$A^{-1} \begin{pmatrix} 1/k \\ \vdots \\ 1/k \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

por tanto

$$\begin{array}{ccccc} 0 & \leq & \frac{1}{k}(\widehat{a}_{11} + \cdots + \widehat{a}_{1k}) & \leq & 1 \\ 0 & \leq & \frac{1}{k}(\widehat{a}_{21} + \cdots + \widehat{a}_{2k}) & \leq & 1 \\ \vdots & & \vdots & & \vdots \\ 0 & \leq & \frac{1}{k}(\widehat{a}_{k1} + \cdots + \widehat{a}_{kk}) & \leq & 1 \end{array}$$

luego, para cada $i = 1, \ldots, k$ se tiene que $0 \leq \sum_{j=1}^{k} \widehat{a}_{ij} \leq k$.

Por otra parte, si para cada $i = 1, \ldots, k$, se verifica que $0 \leq \sum_{j=1}^{k} \widehat{a}_{ij} \leq k$, tenemos que $0 \leq \frac{1}{k} \sum_{j=1}^{k} \widehat{a}_{ij} \leq 1$ para todo $i = 1, \ldots, k$, queda demostrar que $\sum_{i=1}^{k} \left(\frac{1}{k} \sum_{j=1}^{k} \widehat{a}_{ij} \right) = 1$.

Ahora bien

$$\sum_{i=1}^{k} \left(\frac{1}{k} \sum_{j=1}^{k} \widehat{a}_{ij} \right) = \frac{1}{k} \left(\sum_{j=1}^{k} \widehat{a}_{1j} + \cdots + \sum_{j=1}^{k} \widehat{a}_{kj} \right) = \frac{1}{k} \left(\sum_{i=1}^{k} \widehat{a}_{i1} + \cdots + \sum_{i=1}^{k} \widehat{a}_{ik} \right) = \frac{1}{k} \cdot k = 1$$

dándose la penúltima igualdad como consecuencia de la proposición A.1.2, lo que demuestra que $A^{-1}\omega^{(0)} \in \triangle_k$.

3.4.3. $rg(A) = s < k$.

Por el teorema de Weierstrass, existe solución del programa matemático [II]. Si $rg(A) = s < k$ la solución puede no ser única. Además el conjunto de soluciones de un programa convexo como este, es un conjunto convexo lo que permite perfectamente la existencia de infinitas soluciones. A diferencia de cuando A es de rango completo, no está garantizada la compatibilidad de todos los sistemas que se forman en los casos 1 al k y pueden aparecer sistemas compatibles indeterminados cuyas soluciones son de la forma

$$\begin{array}{rcl} \theta_{i_{h+1}} & = & g_1(\theta_{i_1}, \ldots, \theta_{i_h}) \\ \vdots & = & \vdots \\ \theta_{i_k} & = & g_{k-h}(\theta_{i_1}, \ldots, \theta_{i_h}) \end{array}$$

incluidos $\lambda_l = f_l(\theta_{i_1}, \ldots, \theta_{i_h})$ (cuando sean necesarios), debiendo verificar las condiciones de Kuhn-Tucker para convertirse en la solución del programa. Dada la complejidad del proceso, una opción destinada a facilitar el cálculo, es la siguiente:

1. Buscar una solución particular, que denominamos θ_0 que verifique las condiciones de Kuhn-Tucker, en los casos $k - h + 1, \ldots, k$.

2. A partir de esta solución calcular el valor numérico de $\omega^* = A\theta_0$, una vez conocido ω^*, θ_0 se convierte en una solución particular del sistema de ecuaciones lineales $\omega^* = A\theta$. El conjunto de soluciones de este último sistema S^* se puede expresar como $S^* = \{\theta_0\} + S$, siendo S la solución del sistema homogéneo asociado $A\theta = \bar{0}$. La intersección de este subespacio afin S^* con el conjunto factible $\triangle_k$ constituye la solución del programa buscada, lo que se puede expresar como:

$$S_{P.M.} = S^* \cap \triangle_k$$

Es conveniente comentar que los elementos del conjunto $S_{P.M.}$ pueden contener varias componentes nulas y, por otra parte, al ser posible que unas componentes dependan de otras, se produce una pérdida de grados de libertad, estas dos restricciones constituyen un serio inconveniente desde el punto de vista de su interpretación en la práctica.

Capítulo 4

Análisis de Supervivencia

4.1. Análisis de Supervivencia

Las técnicas estadísticas que estudian el tiempo hasta que ocurre un determinado suceso, se engloban dentro de la disciplina de la Estadística que se conoce como Análisis de Supervivencia.

Históricamente, se trató en primer lugar el análisis del tiempo transcurrido hasta la aparición del suceso "muerte". Sin embargo, los métodos estadísticos del Análisis de Supervivencia se aplican igualmente a otros sucesos, que pueden reflejar también el tiempo transcurrido hasta que algo positivo ocurra (por ejemplo, tiempo transcurrido hasta la curación).

Entre los campos principales de aplicación de las técnicas propias del Análisis de Supervivencia cabe destacar: la Ingeniería (donde el Análisis de Supervivencia recibe el nombre de Fiabilidad), la Biomedicina y las Ciencias Sociales. Algunos ejemplos de aplicación a dichas ramas de la Ciencia, pueden ser:

- *Ingeniería*: estudio de la duración de los componentes de un sistema, tiempo hasta que se funde una bombilla, etc.
- *Biomedicina*: estudio del tiempo transcurrido hasta la muerte, curación, remisión de una enfermedad, etc.
- *Ciencias Sociales*: duración del desempleo, duración de los estudios de Licenciatura,

tiempo hasta que se produce el divorcio, la reincidencia, etc.

Para poder analizar el tiempo hasta que ocurre un suceso, hace falta tener definido con claridad el momento que se considera como origen de la observación llamado *instante inicial* y el momento en el que aparece el suceso de interés llamado *punto final.* A partir de estos momentos, la simple resta de estos tiempos proporciona el "tiempo hasta" resultante. En Medicina, el instante inicial suele corresponder al momento en el que el individuo entra en un estudio o experimento, bien porque se le ha diagnosticado una enfermedad, comienza un tratamiento o a la aparición de cualquier otra circunstancia adversa para el individuo (por ejemplo, inicio de exposición a un factor de riesgo). Si el punto final es la muerte del individuo, los datos corresponden literalmente a tiempos de vida o supervivencia, en cualquier otro caso la expresión tiempo de vida tiene un sentido figurado.

La primera referencia a estudios sobre el tiempo de supervivencia a través de datos de mortalidad data del siglo XVIII (ver Hald (1990) y Hosmer y Lemeshow (1999)). Sin embargo, como punto inicial de la aplicación de las técnicas de Análisis de Supervivencia, tal como las entendemos en la actualidad, a las ciencias Biomédicas puede considerarse el trabajo de Berkson y Gage (1950) *Calculation of survival rates for cancer*. En la vertiente paramétrica de comparación de dos poblaciones Cox (1953), en la vertiente no paramétrica para el estudio de supervivencia de una población Kaplan y Meier (1958) y en la vertiente no paramétrica para la comparación de dos poblaciones con Gehan (1965) y Mantel (1966).

Una de las particularidades del Análisis de Supervivencia, debida al hecho de que estudia la variable tiempo, es que los datos no siguen una distribución normal, son asimétricos y son siempre no negativos, y se debe considerar otro tipo de distribuciones: exponenciales, Weibull, Gamma etc, (ver entre otros, Lawless (1982) y Kalbfleisch y Prentice (1980)). Sin embargo, la característica principal del Análisis de Supervivencia es que permite manejar datos censurados o datos con información parcial.

4.1.1. Concepto de censura

Los datos censurados son aquellos que provienen de individuos de los que no se conoce con exactitud su tiempo de supervivencia, bien porque estos hayan abandonado el estudio antes de experimentar el suceso, hayan muerto por causas no relacionadas con el estudio, o simplemente porque el experimento haya terminado sin que hubieran experimentado el suceso. Existen distintos tipos y mecanismos de censura.

Los principales tipos de censura que se suelen considerar son la censura por la derecha por la izquierda y por intervalo.

La *censura por la derecha* se presenta cuando lo único que se sabe acerca de la variable tiempo de supervivencia T es que es mayor que algún valor. Simétricamente al caso anterior, se dice que una variable de tiempo de supervivenvia T está *censurada por la izquierda,* si lo único que se sabe acerca de T es que es menor que algún valor. Por último, la *censura por intervalo,* combina los conceptos de censura por la izquierda y por la derecha, ya que sólo se sabe de T que está entre dos valores.

Básicamente, se pueden distinguir los mecanismos de censura siguiente: censura fija de tipo I, censura fija de tipo II y censura aleatoria.

La censura fija de tipo I se presenta en la situación donde se prefija, por parte del investigador, el tiempo de duración del estudio o periodo de observación t_c. En este caso, en lugar de observar los tiempos de supervivencia $T_1, \ldots, T_n,$ se observan los dados por las variables $Z_1, \ldots, Z_n,$ con $Z_i = T_i$ si $T_i \leq t_c$ y $Z_i = t_c$ si $T_i > t_c$.

En la censura fija de tipo II, el periodo de observación se termina después de haber alcanzado un número prefijado (antes de tomar los datos) de sucesos r. En este caso, en lugar de $T_1, \ldots, T_n$, se observa $Z_1, \ldots, Z_n$, con $Z_{(1)} = T_{(1)}, \ldots, Z_{(r)} = T_{(r)}$, $Z_{(r+1)} = T_{(r)}, \ldots, Z_{(n)} = T_{(r)}$ donde $(\cdot)$ indica el valor ordenado, de menor a mayor de la variable, que ocupa el lugar dado entre paréntesis.

La censura aleatoria se produce cuando se supone que la censura viene dada por una variable aleatoria C independiente de la variable T. Los datos vienen dados según $(Z_i\,,\,\delta_i)$ con $Z_i = \text{mín}\{T_i\,,\,C_i\}$, $\delta_i = 1$ si $T_i \leq C_i$ (dato exacto) y $\delta_i = 0$ si $T_i > C_i$

(dato censurado).

Existen otros mecanismos de censura que intentan responder a las situaciones reales que se analizan, como por ejemplo, censura proporcional en la que se establece una relación entre las variables C y T y que no detallamos aquí, pero sí es importante comentar que despreciar los datos censurados produce una pérdida de información creando, además de sesgos no deseados en las estimaciones, una subjetividad en la eliminación de muestras seleccionadas.

4.1.2. Funciones asociadas al tiempo de supervivencia

El tiempo que transcurre hasta que un suceso ocurre se puede modelizar mediante una variable aleatoria no negativa T. La distribución de T está caracterizada por la función de distribución $F(t)$ y la función de densidad $f(t)$ en el análisis estadístico convencional. En Análisis de supervivencia aparecen, además, asociadas a la variable aleatoria T, otras funciones de interés equivalentes a las anteriores y equivalentes entre sí que caracterizan completamente la distribución y que son la función de supervivencia $S(t)$, la función de riesgo o función tasa de fallo $h(t)$, la función de riesgo acumulado o función de tasa de fallo acumulada $H(t)$ y la función tiempo medio de vida residual $m(t)$.

Función de densidad y de distribución

Los conceptos de función de densidad $f(t)$ y distribución $F(t)$ son los habituales para una variable aleatoria continua, teniendo en cuenta que se trata de funciones definidas para valores no negativos, con lo que

$$f(t) = \lim_{\triangle t \to 0} \frac{P(t \leq T < t + \triangle t)}{\triangle t} \quad \text{con la condición} \quad \int_0^\infty f(t)\, dt = 1$$

La función de distribución se define como $F(t) = P(T \leq t)$ siendo la relación entre ambas

$$F(t) = \int_0^t f(u)\, du$$

Función de Supervivencia

La función de supervivencia se define como $S(t) = P(T > t) = 1 - F(t)$, y representa la probabilidad que tiene un individuo de sobrevivir al instante t, es decir, la probabilidad de experimentar el suceso de interés después del tiempo t. En el contexto industrial, la función $S(t)$ recibe el nombre de función de fiabilidad.

$S(t)$ es una función decreciente con

$$S(0) = 1 \quad \text{y} \quad \lim_{t \to +\infty} S(t) = 0$$

Por tanto, todas las distribuciones teóricas de T tienen siempre la misma forma para $S(t)$, lo que las diferencia es la rapidez con la que $S(t)$ va decreciendo, que depende del "riesgo" asociado a experimentar el suceso y que está medido por otra función $h(t)$, denominada función de riesgo.

Función de riesgo

La función de riesgo $h(t)$ se define como la tasa de muerte (fallo) instantánea para un individuo vivo en el tiempo t, es decir

$$h(t) = \lim_{\triangle t \to 0} \frac{P(\, t \leq T < t + \triangle t \,/\, T \geq t \,)}{\triangle t}$$

donde $P(\, t \leq T < t + \triangle t \,/\, T \geq t \,)$ indica la probabilidad de que un individuo experimente el suceso entre los tiempos t y $t + \triangle t$, sabiendo que ha llegado vivo al tiempo t.

La función de riesgo cuantifica la predisposición al fallo en función del tiempo ya vivido. La lógica de la definición de $h(t)$ está en medir el riesgo instantáneo que tiene un individuo de edad t (que ha llegado vivo a t) de experimentar el suceso.

La función de riesgo puede tener muchas formas (riesgo creciente, decreciente, constante, tipo bañera o "bath-tube",...) y presenta las siguientes propiedades:

a) $h(t) \geq 0$ para todo $t \in [0, \infty)$

b) $\lim_{t \to +\infty} \int_0^t h(t)\, dt = \infty$

El producto $h(t) \cdot \triangle t$ se puede considerar como la probabilidad aproximada que tiene

un individuo de edad t de experimentar el suceso en el instante siguiente. Pero hay que tener en cuenta, sin embargo, que $h(t)$ no es una probabilidad.

A la función $h(t)$ se la conoce con distintos nombres dependiendo del campo de aplicación en el que se esté, así en Fiabilidad es la tasa de fallo condicional ("conditional failure rate"), en Demografía es la fuerza de la mortalidad ("force of mortality"), en Procesos Estocásticos es la función de intensidad ("intensity function"). Pero en la mayoría de ocasiones es conocida como función de riesgo.

Por otra parte, se puede demostrar que

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} Ln\, S(t) \quad \text{y que} \quad h(t) = \frac{-S'(t)}{S(t)}$$

como vemos a continuación:

$$P(t \leq T < t + \triangle t \,/\, T \geq t) = \frac{P(t \leq T < t + \triangle t)}{P(T \geq t)} = \frac{F(t + \triangle t) - F(t)}{S(t)},$$

por lo que

$$h(t) = \lim_{\triangle t \to 0} \left(\frac{1}{\triangle t} \frac{F(t + \triangle t) - F(t)}{S(t)} \right) = \frac{1}{S(t)} \lim_{\triangle t \to 0} \left(\frac{F(t + \triangle t) - F(t)}{\triangle t} \right) = \frac{F'(t)}{S(t)} = \frac{f(t)}{S(t)}.$$

dándose el resto de las relaciones como consecuencia del penúltimo cociente.

Función de riesgo acumulada

La función de riesgo acumulada $H(t)$ se define como

$$H(t) = \int_0^t h(x)dx$$

y también se la conoce como función tasa de fallo acumulada. Esta función verifica que $H(t) = -Ln\, S(t)$ y por tanto

$$S(t) = exp(-H(t)),$$

La función de riesgo acumulado verifica las siguientes propiedades:

a) $H(t_1) \leq H(t_2)$ si $t_1 < t_2$

b) $\lim_{t\to 0} H(t) = 0$ y $\lim_{t\to\infty} H(t) = \infty$.

Función tiempo medio de vida residual

La función tiempo medio de vida residual ("mean residual lifetime") $mrl(t)$ ó $m(t)$ se define como

$$m(t) = E[T - t \,/\, T > t]$$

y representa la esperanza de vida para un individuo que haya sobrevivido t unidades de tiempo.

4.1.3. Relaciones entre las funciones teóricas de supervivencia

Anteriormente se han definido distintas funciones que aparecen en el Análisis de Supervivencia. Dichas funciones se relacionan entre sí de una forma cíclica y es posible conocer, a partir de una de ellas, las restantes.

Para el ciclo $H \to S \to mrl \to h \to f \to F \to H$ las relaciones matemáticas son:

$$\begin{aligned}
S(t) &= exp[-H(t)] \\
mrl(t) &= \frac{\int_t^\infty S(y)\,dy}{S(t)} \\
h(t) &= \frac{\frac{d}{dt}\,mrl(t) + 1}{mrl(t)} \\
f(t) &= h(t)\,exp\left[-\int_0^t h(y)\,dy\right] \\
F(t) &= \int_0^t f(y)dy \\
H(t) &= -Ln[1 - F(t)]
\end{aligned}$$

Para el ciclo $H \leftarrow S \leftarrow mrl \leftarrow h \leftarrow f \leftarrow F \leftarrow H$ las relaciones matemáticas son:

$$
\begin{aligned}
H(t) &= -LnS(t) \\
S(t) &= \frac{mrl(0)}{mrl(t)}\, exp\left[-\int_0^t \frac{dy}{mrl(y)}\right] \\
mrl(t) &= \frac{\displaystyle\int_t^\infty exp\left[-\int_0^y h(x)\,dx\right] dy}{\displaystyle exp\left[-\int_0^t h(y)dy\right]} \\
h(t) &= \frac{f(t)}{\displaystyle 1-\int_0^t f(y)\,dy} \\
f(t) &= F'(t) \\
F(t) &= 1 - exp[-H(t)].
\end{aligned}
$$

4.2. Modelos paramétricos

Hablamos de modelos paramétricos cuando suponemos que la distribución teórica $F(t)$ de la variable aleatoria T, pertenece a una familia $\mathcal{F}$ formada por distribuciones de forma funcional fija y conocida, dependientes de uno o más parámetros reales. Numerosos modelos paramétricos se han utilizado en análisis de supervivencia. En particular y debido a su utilidad en un amplio número de situaciones destacamos las distribuciones: Exponencial, Weibull, Valor extremo, Gamma.

Distribución Exponencial

Históricamente la distribución exponencial fue la primera que se utilizó de forma generalizada como distribución de tiempos de vida debido, en parte, a la simplicidad de los métodos estadísticos de los que se disponía y, en parte, también a que la distribución representaba bien los tiempos de vida de bastantes productos manufacturados, Davis (1952), Epstein (1958). En medicina se ha utilizado en el estudio de la supervivencia o remisión de enfermedades crónicas Feigl y Zelen (1965). Su característica más importante es que su función de riesgo es constante en el tiempo, lo que ha llevado a que se le conozca, como distribución *sin memoria*, expresión que resume la idea de que el riesgo no depende del tiempo transcurrido.

Se dice que la variable aleatoria T tiene distribución Exponencial de parámetro $\lambda > 0$, que denotamos como $T \sim Exp(\lambda)$ si su función de densidad es

$$f(t) = \lambda e^{-\lambda t} \qquad t \geq 0$$

siendo su función de supervivencia

$$S(t) = e^{-\lambda t} \qquad t \geq 0$$

y por tanto la función de riesgo es

$$h(t) = \lambda.$$

Si tomamos $\theta = \lambda^{-1}$ entonces la función de densidad es:

$$f(t) = \frac{1}{\theta} e^{-t/\theta} \qquad t \geq 0$$

siendo la media y la varianza de la distribución θ y θ^2 respectivamente. Cuando $\theta = 1$ decimos que se trata de la distribución exponencial estándar.

- Distribución Weibull

La distribución Weibull, es sin duda, la más extendida y utilizada de las distribuciones de tiempos de vida, debe su nombre a Waloddi Weibull (1951). Se trata de un modelo de alta flexibilidad debido a su gran variedad de formas lo que le permite adaptarse bien a distintos tipos de datos, hecho que unido a la sencillez de su expresión matemática ha propiciado su popularidad. La distribución Weibull se ha utilizado en distintas ramas de la ingeniería, ver por ejemplo, Kao (1959), Lieblein y Zelen (1956) y de las ciencias biomédicas, ver por ejemplo Peto y Lee (1973), Whittemore y Altschuler (1976).

Decimos que la variable aleatoria T tiene distribución Weibull de parámetros $\lambda > 0$, $\beta > 0$, que denotamos como $T \sim W(\lambda, \beta)$ si su función de densidad es

$$f(t) = \lambda\beta(\lambda t)^{\beta-1}\, exp[-(\lambda t)^{\beta}], \qquad t > 0$$

La función de supervivencia viene dada por:

$$S(t) = exp[-(\lambda t)^{\beta}], \qquad t > 0$$

y por tanto, la función de riesgo es:

$$h(t) = \lambda\beta(\lambda t)^{\beta-1}, \qquad t \geq 0$$

la función de riesgo es creciente si $\beta > 1$, decreciente si $\beta < 1$ y constante si $\beta = 1$. La media y la varianza de esta distribución, son:

$$\lambda^{-1}\Gamma(1 + 1/\beta) \qquad \text{y} \qquad \lambda^{-2}[\Gamma(1 + 2/\beta) - \Gamma(1 + 1/\beta)^2]$$

respectivamente y en general $E[T^r] = \lambda^{-r}\Gamma(1 + r/\beta)$, siendo Γ la función Gamma.

La forma de la distribución Weibull, depende del parámetro β, conocido como parámetro de forma de la distribución, estando sus valores generalmente comprendidos entre 0.5 y 3. El otro parámetro λ es un parámetro de escala. La distribución Weibull incluye como caso particular $(\beta = 1)$ a la distribución exponencial.

- Distribución Valor extremo

La distribución valor extremo, conocida también como distribución Gumbel, puesto que fue E. J. Gumbel (1958) quien comenzó a utilizarla, describe adecuadamente algunos tipos de fenómenos de carácter físico, tales como precipitaciones durante periodos de sequía, resistencia eléctrica, etc. y también, ciertos tiempos de vida como, por ejemplo, la mortalidad humana debida a la edad. Su función de densidad es:

$$f(x) = \frac{1}{b}\, exp\left[\frac{x-u}{b} - exp\left(\frac{x-u}{b}\right)\right] \qquad -\infty < x < \infty$$

la de supervivencia

$$S(x) = exp\left[-exp\left(\frac{x-u}{b}\right)\right] \qquad -\infty < x < \infty$$

y la de riesgo

$$h(x) = \frac{1}{b}\, exp\left[\left(\frac{x-u}{b}\right)\right] \qquad -\infty < x < \infty$$

siendo $b > 0$ y u $(-\infty < u < \infty)$ los paramétros.

Esta distribución está directamente relacionada con la distribución Weibull ya que si la variable aleatoria T tiene distribución Weibull de parámetros (λ, β), la variable aleatoria $X = log\, T$ tiene distribución valor extremo con parámetros $b = \beta^{-1}$ y $u = -log\lambda$.

La distribución valor extremo con $u = 0$ y $b = 1$ es conocida como distribución valor extremo estándar, estando tabulada por Meeker y Nelson (1974), y sus momentos de orden uno y dos son:

$$\int_{-\infty}^{\infty} x\, exp(x - e^x)dx = -\gamma$$
$$\int_{-\infty}^{\infty} x^2\, exp(x - e^x)dx = \frac{\pi^2}{6} + \gamma^2$$

y su varianza $\pi^2/6$, con $\gamma = 0.5772...$ la constante de Euler. Para cualquier otra distribución con parámetros de localización u y escala b, la media es $u - \gamma b$ y la varianza $(\pi^2/6)b^2$.

- Distribución Gamma

La distribución gamma, ha sido utilizada como distribución de tiempos de vida, (ver por ejemplo Gupta y Groll (1961)) y de otras variables aleatorias no negativas. Se dice que la variable aleatoria T sigue una distribución gamma de parámetros λ y k, $T \sim G(\lambda, k)$, si su función densidad es

$$f(t) = \frac{\lambda^k}{\Gamma(k)}\, t^{k-1}e^{-\lambda t} \qquad t > 0$$

siendo $\lambda > 0$ el parámetro de escala, $k > 0$ el parámetro de forma y Γ la función gamma.

La función de distribución es

$$F(t) = \int_0^t \frac{\lambda^k}{\Gamma(k)}\, u^{k-1}e^{-\lambda u}du \qquad t > 0$$

y aunque cuando k es un número entero positivo (distribución Erlang) se conoce la primitiva de esta integral, en general sus probabilidades se calculan utilizando tablas para distintos valores de los parámetros. La función de riesgo $h(t) = f(t)/S(t)$ es creciente para $k > 1$ y decreciente para $0 < k < 1$.

La media y la varianza de la distribución son

$$E[T] = \frac{\lambda}{k}, \qquad var[T] = \frac{\lambda}{k^2}$$

y el momento de orden r viene dado por $E[T^r] = \dfrac{\Gamma(k+r)}{\lambda^r\Gamma(k)}$

La distribución $G\left(\frac{1}{2}, \frac{m}{2}\right)$ es conocida como distribución ji-cuadrado χ^2_m con m grados de libertad.

La distribución exponencial de parámetro λ aparece como caso particular de la distribución gamma cuando $k = 1$, además, si $T_1, \ldots, T_n$ son variables aleatorias independientes con distribución exponencial de parámetro λ, entonces la variable aleatoria $S = T_1 + \cdots + T_n$ sigue una distribución $G(\lambda, n)$.

4.3. Modelos no paramétricos

A diferencia del apartado anterior, ahora consideramos que la familia $\mathcal{F}$ a la que pertenece la distribución teórica $F(t)$ de la variable aleatoria T es no paramétrica, por tanto queda abierto un amplio abanico de posibilidades para $\mathcal{F}$ que, por ejemplo, puede contener todas las funciones de distribución continuas, absolutamente continuas, etc; generalmente se supone la diferenciabilidad de F.

En muchas situaciones reales el punto de partida se sitúa entre los dos tipos de modelos (paramétricos y no paramétricos) y ello se debe a la información disponible acerca de la función $F(t)$, por lo cual se prefiere un modelo que incorpore dicha información. En general, la información adicional sobre $F(t)$ se formula en términos de un conjunto de restricciones de información que son usualmente restricciones de momento (Zellner y Highfield, 1988).

En análisis de supervivencia hay ocasiones en las que la función de riesgo $h_F(t)$ debe satisfacer ciertas restricciones, por ejemplo $h_F(t)$ es una función creciente de t ó $h_F(t) = \theta$ con $\theta > 0$ o también $[h_F(t)]^{-1}$ es una función cóncava. La diferenciabilidad de la densidad $f(t)$ también puede ser un punto de partida (Ebrahimi, 2000).

Cuando se trata de inferir la distribución F en una de estas situaciones, uno de los métodos "no paramétricos" de inferencia es el basado en el principio de máxima entropía. Pues bien, Ebrahimi (2000) muestra como usando este principio, en distintos supuestos no paramétricos con restricciones de información sobre la función de riesgo, se obtienen como distribuciones estimadas modelos paramétricos como el exponencial, Pareto, valor extremo y otros. Estas aproximaciones nos parece que ilustran las relaciones entre las

modelizaciones paramétrica y no paramétrica.

4.4. Modelos de supervivencia discretos

Todos los modelos descritos, tanto en el apartado de modelos paramétricos como los aludidos en el de no paramétricos, resultan apropiados para datos de supervivencia provenientes de distribuciones de probabilidad continuas. Sin embargo, en ocasiones los datos de supervivencia son discretos, bien debido al agrupamiento de observaciones de datos continuos por la imprecisión de la medida, bien debido a la propia naturaleza del tiempo medido.

Cualquiera de los modelos paramétricos descritos puede generar un modelo discreto introduciendo un agrupamiento en el eje T. Por ejemplo, si el tiempo de vida sigue una distribución Weibull con función de supervivencia

$$S(t) = exp[-(\lambda t)^{\beta}], \qquad t > 0$$

y los tiempos se agrupan en intervalos de amplitud unidad, de forma que la variable discreta observada es $T_1 = [T]$, donde $[T]$ representa "la parte entera de T", la función de probabilidad de T_1 puede escribirse como:

$$\begin{aligned} p(t_1) &= P[T_1 = t_1] = P(t_1 \leq T < t_1 + 1) = \\ &= \theta^{t_1^{\beta}} - \theta^{(t_1+1)^{\beta}} \qquad t_1 = 0, 1, 2, \ldots \qquad [*] \end{aligned}$$

siendo $0 < \theta = exp(-\lambda^{\beta}) < 1$. El caso especial $\beta = 1$ es la distribución geométrica con función de probabilidad $\theta^{t_1}(1-\theta)$. La función de riesgo correspondiente a $[*]$ es

$$h(t_1) = P(T_1 = t_1 / T_1 \leq t_1) = 1 - \theta^{(t_1+1)^{\beta} - t_1^{\beta}}$$

que es monótona creciente, monótona decreciente, ó constante para $\beta > 1$, $\beta < 1$ o $\beta = 1$ respectivamente.

4.4.1. Modelo de supervivencia no paramétrico con datos agrupados

Sea T la variable aleatoria no negativa que representa el tiempo de vida en un estudio de supervivencia, siendo F su función de distribución, que supondremos absolutamente continua.

Consideremos una partición del tiempo

$$(0, \infty) = \bigcup_{i=1}^{k} (t_{i-1},\, t_i] \quad \text{con} \quad t_0 = 0 \quad \text{y} \quad t_k = \infty.$$

Por restricciones de observación, el investigador sólo puede observar la variable en los instantes $t_1 \leq t_2 \leq \cdots \leq t_{k-1}$ (al final de cada hora, día o periodo similar, no necesariamente de igual duración), de forma que los tiempos de vida de las unidades muestrales se registran agrupados en los k intervalos $(t_{i-1},\, t_i] \quad i = 1, \ldots, k.$

Para cada $1 \leq i \leq k-1,$ definimos.

$$\theta_i = \int_{t_{i-1}}^{t_i} f(t)\, dt, \qquad \theta_k = 1 - \sum_{i=1}^{k-1} \theta_i$$

θ_i representa la probabilidad de morir en $(t_{i-1},\, t_i]\,,$ θ_k representa la probabilidad de sobrevivir al instante t_{k-1}. Estamos ante un modelo de supervivencia con variable tiempo de vida discretizada por restricciones de observación.

La variable observada para cada unidad experimental que denotamos por d_0 presenta k modalidades excluyentes (morir en cualquiera de los k intervalos de tiempo) y podrá expresarse mediante una variable aleatoria discreta k-dimensional con distribución multinomial:

$$d_0 = (x_1, \ldots, x_k) \equiv Mu(1, \theta).$$

de parámetros 1 y vector de probabilidades θ con componentes $\theta_i, \quad i = 1, \ldots, k.$

4.4.2. Modelo de supervivencia no paramétrico censurado aleatoriamente por la derecha y datos agrupados

Sea T la variable aleatoria no negativa que representa el tiempo de vida en un estudio de supervivencia, siendo F su función de distribución, que supondremos absolutamente continua.

Consideremos una partición del tiempo

$$(0,\infty) = \bigcup_{i=1}^{k} (t_{i-1},\, t_i] \quad \text{con} \quad t_0 = 0 \quad \text{y} \quad t_k = \infty.$$

Por restricciones de observación, el investigador sólo puede observar la variable en los instantes $t_1 \leq t_2 \leq \cdots \leq t_{k-1}$ (al final de cada hora, día o periodo similar, no necesariamente de igual duración), de forma que los tiempos de vida de las unidades muestrales se registran agrupados en los k intervalos $(t_{i-1},\, t_i]$ $i = 1, \ldots, k$.

Además, en estos instantes t_i, la variable se censura aleatoriamente según una distribución de probabilidad conocida. Denotamos por C a la variable de censura. De esta forma la información que recoge el investigador para cada elemento de la muestra es o bien el intervalo $(t_{i-1},\, t_i]$ en el que éste muere (observación no censurada) o bien el instante t_i al que sobrevive (observación censurada). Supondremos que un tiempo de vida censurado en el instante t_i es superior a dicho tiempo.

Para cada $1 \leq i \leq k-1$, definimos.

$$\theta_i = \int_{t_{i-1}}^{t_i} f(t)\, dt, \qquad \theta_k = 1 - \sum_{i=1}^{k-1} \theta_i$$

θ_i representa la probabilidad de morir en $(t_{i-1},\, t_i]$, θ_k representa la probabilidad de sobrevivir al instante t_{k-1}.

$$c_i = P(C = t_i), \quad i = 1, \ldots k-1 \quad \text{y} \quad c_k = 1 - \sum_{i=1}^{k-1} c_i$$

c_i representa la probabilidad de censurar en t_i, $i = 1, \ldots, k-1$ y c_k la probabilidad de no censurar. Supongamos además que $c_{k-1} + c_k > 0$, restricción necesaria para poder

observar las unidades muestrales más allá del instante t_{k-2}.

La siguiente figura, muestra esquemáticamente las probabilidades que maneja el modelo y el periodo o instante al que se refieren.

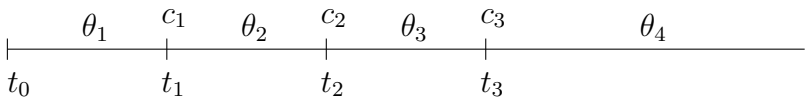

Figura 4.1: Representación del modelo para $k = 4$

La variable observada para cada individuo de la muestra, que denotamos por d_c, presenta $2k-1$ modalidades excluyentes (morir en cualquiera de los k intervalos de tiempo o censurar en cualquiera de los $k-1$ instantes de tiempo) con lo que podemos expresarla mediante una variable aleatoria discreta $(2k-1)$-dimensional con distribución multinomial

$$d_c = (x_1, \ldots, x_k,\, y_1, \ldots, y_{k-1}) \equiv Mu(1;\, \omega(\theta))$$

con

$$\omega(\theta) = \left(\theta_1 \sum_{j=1}^{k} c_j,\, \theta_2 \sum_{j=2}^{k} c_j, \ldots, \theta_k c_k,\, c_1 \sum_{j=2}^{k} \theta_j,\, c_2 \sum_{j=3}^{k} \theta_j, \ldots, c_{k-1}\theta_k\right)$$

y que se puede escribir para reducir la notación como

$$\omega(\theta) = (\theta_1 c_{(1)}, \ldots, \theta_k c_k,\, \theta_{(2)} c_1, \ldots, \theta_k c_{k-1})$$

siendo $c_{(i)} = \sum_{j=i}^{k} c_j, \quad \theta_{(i)} = \sum_{j=i}^{k} \theta_j \qquad i = 1, \ldots, k-1.$

Este modelo se convierte en el modelo visto en 4.4.1 cuando las variables $Y_1, \ldots, Y_k$ del vector d_c resultan degeneradas en 0, es decir, en ausencia de censura, $c_i = 0$, $i = 1, \ldots, k-1$; $c_k = 1$.

El modelo anterior fue presentado en Turrero (1988) y ha sido estudiado en un contexto Bayesiano en Turrero (1989) y en el contexto de medidas de información paramétricas en Turrero (1995).

Capítulo 5

Aplicación a un modelo de Supervivencia

Consideremos el modelo definido en 4.4.2 y que resumimos a continuación:

- Sea $T \geq 0$ una variable aleatoria, con función de densidad desconocida
- Sea $\{(t_{i-1},\, t_i\,]\}$, $i = 1, \ldots, k$, una partición de $(0, \infty)$ con $t_0 = 0$ y $t_k = \infty$.
- Sea C una variable aleatoria discreta que representa el tiempo de censura, siendo $\{t_1, \ldots, t_{k-1}\}$ su soporte.

  Para cada unidad experimental sólo se puede observar el intervalo $(t_{i-1},\, t_i]$ $i = 1, \ldots, k$ donde "muere", o el instante t_i donde se censura.
- Para cada $1 \leq i \leq k-1$,

$$\theta_i = \int_{t_{i-1}}^{t_i} f(t)\, dt, \qquad \theta_k = 1 - \sum_{i=1}^{k-1} \theta_i$$

$$c_i = P(C = t_i) \qquad c_k = 1 - \sum_{i=1}^{k-1} c_i$$

  θ_i representa la probabilidad de morir en $(t_{i-1},\, t_i]$ y c_i la probabilidad de censurar en t_i $(i = 1, \ldots, k-1)$, θ_k representa la probabilidad de sobrevivir al instante t_{k-1}, y c_k la probabilidad de no censurar.
- Se supone que todas las c_i son conocidas y que T y C son independientes.

Una vez fijada la distribución de C, que se denota por $c = (c_1, c_2, \ldots, c_k)$ con $c_i \geq 0$ $(i = 1, \ldots, k)$ y $c_{k-1} + c_k > 0$, se genera el experimento ϵ_c, que consiste en la observación de la variable $(2k-1)$-dimensional:

$$d_c = (x_1, \ldots, x_k\,,\, y_1, \ldots, y_{k-1}) \equiv M_u(1; \omega(\theta)), \text{ con}$$

$$\omega(\theta) = (\theta_1 c_{(1)}, \ldots, \theta_k c_k, \theta_{(2)} c_1, \ldots, \theta_k c_{k-1})$$

siendo

$$c_{(i)} = \sum_{j=i}^{k} c_j \qquad \theta_{(i)} = \sum_{j=i}^{k} \theta_j \quad (i = 1, \ldots, k-1)$$

Se denota por ϵ_0 al experimento ϵ_c cuando $c = (0, \ldots, 0, 1)$, es decir el experimento no censurado. Ahora las variables $y_1, \ldots, y_{k-1}$ del vector d_c son degeneradas en 0 y la variable observada es la variable k-dimensional:

$$d_0 = (x_1, \ldots, x_k) \equiv Mu(1, \theta).$$

$\omega(\theta)$ se puede expresar mediente la siguiente ecuación matricial

$$\omega = A\theta$$

siendo A la matriz de orden $(2k-1) \times k$ siguiente:

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & c_{(2)} & 0 & \cdots & 0 & 0 \\ 0 & 0 & c_{(3)} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_{(k-1)} & 0 \\ 0 & 0 & 0 & \cdots & 0 & c_k \\ 0 & c_1 & c_1 & \cdots & c_1 & c_1 \\ 0 & 0 & c_2 & \cdots & c_2 & c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_{k-2} & c_{k-2} \\ 0 & 0 & 0 & \cdots & 0 & c_{k-1} \end{pmatrix}$$

y $\theta \in \triangle_k$, por tanto constiuye un caso particular de los estudiados en el capítulo III, siendo aquí $r = 2k - 1$, (se puede observar que la suma de los elementos de cada una de las columnas de la matriz A es 1).

Al igual que en el capítulo III, se quiere estudiar el comportamiento de $H \in \mathcal{H}$, como función de θ, $H(\omega(\theta)) = H(A\,\theta)$ más concretamente, caracterizar el valor o valores de θ que maximizan dicha entropía.

5.1. Formulación del programa

El problema de programación matemática que debemos resolver es el siguiente:

$$\min \| \omega^{(0)} - A\,\theta \|^2$$

sujeto a las restricciones

$$(i) \qquad \sum_{j=1}^{k} \theta_j = 1$$

$$(ii) \qquad \theta_j \geq 0 \quad j = 1, \ldots, k$$

Como ya se ha visto en el capítulo III, la función objetivo $G(\theta)$ es continua, diferenciable y convexa en $\mathbb{R}^k$ como función de θ, y el conjunto de soluciones factible $\triangle_k$ es cerrado, acotado y convexo, se trata, pues, de un programa convexo para mínimo. Solamente nos queda por analizar en qué situaciones A es de rango completo y por tanto, la función objetivo estrictamente convexa.

Estudio del rango de la matriz A.

Por ser $c_{k-1} + c_k > 0$ la matriz A es de rango completo para todo $c = (c_1, \ldots, c_k)$ ya que el menor de orden $k \times k$ siguiente es distinto de 0

$$\begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & c_{(2)} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & c_{(3)} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_{(k-2)} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & c_{(k-1)} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & {}^{*}c_{k-1}(c_k) \end{vmatrix} \neq 0$$

* Elegir entre c_{k-1} ó c_k el que sea distinto de 0 o cualquiera de ellos si ambos son distintos de 0.

5.1.1. Resolución del programa

Por el teorema de Weierstrass sabemos que existe solución del programa y además sabemos que va a ser única, pues la función objetivo es estrictamente convexa (en cualquier situación). Para encontrar la solución, utilizamos las técnicas de programación matemática descritas en los capítulos anteriores, en concreto el Teorema de Kuhn-Tucker. Las condiciones necesarias y suficientes de Kuhn-Tucker que debe verificar un punto θ^* candidato a óptimo (en este caso global) son las siguientes:

$$(1) \quad \frac{\partial L}{\partial \theta_j}(\theta^*) = 0 \qquad j = 1, \ldots, k$$

$$(2) \quad \lambda_j \, \theta_j^* = 0 \qquad j = 1, \ldots, k$$

$$(3) \quad \theta_j^* \geq 0 \qquad j = 1, \ldots, k$$

$$(4) \quad \lambda_j \leq 0 \qquad j = 1, \ldots, k$$

$$(5) \quad \sum_{j=1}^{k} \theta_j^* = 1$$

siendo L la función Lagrangiana:

$$L = \left(\frac{1}{2k-1} - \theta_1\right)^2 + \left(\frac{1}{2k-1} - \theta_{(2)}c_1\right)^2 + \cdots + \left(\frac{1}{2k-1} - \theta_{k-1}c_{(k-1)}\right)^2 +$$

$$+ \left(\frac{1}{2k-1} - \theta_k c_{k-1}\right)^2 + \left(\frac{1}{2k-1} - \theta_k c_k\right)^2 + \sum_{j=1}^{k} \lambda_j \theta_j + \mu \left(\sum_{j=1}^{k} \theta_j - 1\right)$$

Para resolver analíticamente el programa, hay que encontrar las soluciones del conjunto de ecuaciones formado por las condiciones de Kuhn-Tucker (1) a (5) anteriores, es decir, hay un total de $4k+1$ condiciones.

Las hipótesis que podemos hacer sobre los valores que toman los $\lambda_j,\ j=1,\ldots,k,$ que dan lugar a los casos $1,\ldots,k,$ ya han sido analizadas detalladamente en el capítulo III. Seguidamente demostraremos que la solución óptima que buscamos se encuentra siempre en el caso 1.

Caso 1. $\lambda_1=\lambda_2=\cdots=\lambda_k=0$. La función Lagrangiana que queda es

$$\begin{aligned} L &= \left(\frac{1}{2k-1}-\theta_1\right)^2+\left(\frac{1}{2k-1}-\theta_{(2)}c_1\right)^2+\cdots+\left(\frac{1}{2k-1}-\theta_{k-1}c_{(k-1)}\right)^2+ \\ &\quad +\left(\frac{1}{2k-1}-\theta_k c_{k-1}\right)^2+\left(\frac{1}{2k-1}-\theta_k c_k\right)^2+\mu(\theta_1+\cdots+\theta_k-1) \end{aligned}$$

Las derivadas parciales de la función L, llamando $s=1/(2k-1),$ son

$$\frac{\partial L}{\partial\theta_1} = 2\left(s-\theta_1\right)(-1)+\mu$$

$$\frac{\partial L}{\partial\theta_2} = 2\left(s-\theta_{(2)}c_1\right)(-c_1)+2\left(s-\theta_2 c_{(2)}\right)(-c_{(2)})+\mu$$

para $2<j\leq k-1$

$$\begin{aligned} \frac{\partial L}{\partial\theta_j} &= 2\left(s-\theta_{(2)}c_1\right)(-c_1)+2\left(s-\theta_{(3)}c_2\right)(-c_2)+\cdots+2\left(s-\theta_{(j)}c_{j-1}\right)(-c_{j-1})+ \\ &\quad +2\left(s-\theta_j c_{(j)}\right)(-c_{(j)})+\mu \end{aligned}$$

y cuando $j=k$

$$\begin{aligned} \frac{\partial L}{\partial\theta_k} &= 2\left(s-\theta_{(2)}c_1\right)(-c_1)+\cdots+2\left(s-\theta_{(k-1)}c_{k-2}\right)(-c_{k-2})+2\left(s-\theta_k c_{k-1}\right)(-c_{k-1})+ \\ &\quad +2\left(s-\theta_k c_k\right)(-c_k)+\mu \end{aligned}$$

que al igualarlas a 0 forman el sistema:

$$\left.\begin{array}{rcl}
2\left(s-\theta_1\right)\left(-1\right)+\mu & = & 0 \\
2\left(s-\theta_{(2)}c_1\right)\left(-c_1\right)+2\left(s-\theta_2 c_{(2)}\right)\left(-c_{(2)}\right)+\mu & = & 0 \\
\vdots \qquad\qquad\qquad \vdots & & \vdots \\
2\left(s-\theta_{(2)}c_1\right)\left(-c_1\right)+\cdots+2\left(s-\theta_{(k-1)}c_{k-2}\right)\left(-c_{k-2}\right)+2\left(s-\theta_{k-1}c_{(k-1)}\right)\left(-c_{(k-1)}\right)+\mu & = & 0 \\
2\left(s-\theta_{(2)}c_1\right)\left(-c_1\right)+\cdots+2\left(s-\theta_k c_{k-1}\right)\left(-c_{k-1}\right)+2\left(s-\theta_k c_k\right)\left(-c_k\right)+\mu & = & 0
\end{array}\right\}$$

Para resolverlo, teniendo en cuenta además que nuestro interés es demostrar también que $\theta_j > 0,\ \ j = 1, \ldots, k$ (caso 1), procedemos de forma diferente a la vista en el capítulo III, lo que nos va a permitir demostrar de manera más sencilla que, efectivamente, la solución buscada siempre se encuentra dentro del caso 1.

Igualando $\dfrac{\partial L}{\partial \theta_{k-1}} = \dfrac{\partial L}{\partial \theta_k}$ queda después de simplificar

$$2\left(\frac{1}{2k-1}-\theta_{k-1}c_{(k-1)}\right)\left(-c_{(k-1)}\right) = 2\left(\frac{1}{2k-1}-\theta_k c_{k-1}\right)\left(-c_{k-1}\right)+2\left(\frac{1}{2k-1}-\theta_k c_k\right)\left(-c_k\right)$$

por tanto

$$\begin{aligned}
\theta_{k-1}c_{(k-1)}^2 &= \theta_k c_{k-1}^2 + \theta_k c_k^2 \\
\theta_{k-1} &= \theta_k\left(\frac{c_{k-1}^2+c_k^2}{c_{(k-1)}^2}\right) \\
\theta_{k-1} &= \theta_k A_{k-1}, \quad \text{siendo } A_{k-1} = \left(\frac{c_{k-1}^2+c_k^2}{c_{(k-1)}^2}\right)
\end{aligned}$$

Igualando $\dfrac{\partial L}{\partial \theta_{j+1}} = \dfrac{\partial L}{\partial \theta_j}$ para $j = 1, \ldots, k-2$ queda:

$$\theta_j c_{(j)}^2 = \left[\theta_{j+1}+\theta_{j+2}+\cdots+\theta_k\right]c_j^2 + \theta_{j+1}c_{(j+1)}^2 \qquad (1)$$

en particular, para $j = k-2$ se tiene

$$\begin{aligned}
\theta_{k-2}c^2_{(k-2)} &= [\theta_{k-1}+\theta_k]c^2_{k-2}+\theta_{k-1}c^2_{(k-1)} \\
\theta_{k-2}c^2_{(k-2)} &= [\theta_k A_{k-1}+\theta_k]c^2_{k-2}+\theta_k A_{k-1}c^2_{(k-1)} \\
\theta_{k-2} &= \theta_k\left(\frac{[1+A_{k-1}]c^2_{k-2}+A_{k-1}c^2_{(k-1)}}{c^2_{(k-2)}}\right) \\
\theta_{k-2} &= \theta_k A_{k-2}
\end{aligned}$$

procediendo de la misma forma para $j=k-3,k-2,\ldots$ resulta que es posible expresar cada $\theta_j,\ j=1,\ldots,k-1$ como el producto de θ_k por un factor que denominamos $A_j,\ j=1,\ldots,k-1$ es decir,

$$\theta_j=\theta_k A_j, \qquad j=1,\ldots,k-1.$$

A partir de la ecuación (1) se puede obtener una expresión general para A_j ya que

$$\begin{aligned}
\theta_j c^2_{(j)} &= \left[\theta_{j+1}+\theta_{j+2}+\cdots+\theta_k\right]c^2_j+\theta_{j+1}c^2_{(j+1)} \\
\theta_j c^2_{(j)} &= \theta_k\left[A_{j+1}+A_{j+2}+\cdots+A_{k-1}+1\right]c^2_j+\theta_k A_{j+1}c^2_{(j+1)} \\
\theta_j &= \theta_k\cdot\frac{\left[(A_{j+1}+\cdots+A_{k-1}+1)c^2_j+A_{j+1}c^2_{(j+1)}\right]}{c^2_{(j)}}
\end{aligned}$$

por tanto:

$$\boxed{\begin{aligned}
A_j &= \frac{\left(1+\sum_{l=j+1}^{k-1}A_l\right)c^2_j+A_{j+1}c^2_{(j+1)}}{c^2_{(j)}}, \qquad j=1,\ldots,k-2 \\
A_{k-1} &= \frac{c^2_{k-1}+c^2_k}{c^2_{(k-1)}}
\end{aligned}}$$

Para demostrar que $\theta_j>0$ para todo $j=1,\ldots,k$ descomponemos A_j de la siguiente forma:

$$A_j=\frac{\left(1+\sum_{l=j+1}^{k-1}A_l\right)c^2_j}{c^2_{(j)}}+\frac{A_{j+1}c^2_{(j+1)}}{c^2_{(j)}}, \quad j=1,\ldots,k-2$$

al ser $A_{k-1} > 0$, se observa por recurrencia que

$$\left.\begin{array}{rl} \dfrac{\left(1+\displaystyle\sum_{l=j+1}^{k-1} A_l\right) c_j^2}{c_{(j)}^2} & \geq 0 \\ \dfrac{A_{j+1} c_{(j+1)}^2}{c_{(j)}^2} & > 0 \end{array}\right\} \Rightarrow A_j > 0 \quad \text{para todo} \quad j = 1, \ldots, k-2.$$

De $\theta_1 + \cdots + \theta_k = 1$ se obtiene sustituyendo:

$$\begin{aligned} A_1\theta_k + \cdots + A_{k-1}\theta_k + \theta_k &= 1 \\ \theta_k(A_1 + \cdots + A_{k-1} + 1) &= 1 \end{aligned}$$

por tanto

$$\begin{aligned} \theta_k^* &= \frac{1}{A_1 + \cdots + A_{k-1} + 1} \\ \theta_{k-1}^* &= \frac{A_{k-1}}{A_1 + \cdots + A_{k-1} + 1} \\ &\vdots \qquad\qquad \vdots \\ \theta_1^* &= \frac{A_1}{A_1 + \cdots + A_{k-1} + 1} \end{aligned}$$

luego $\theta_j^* > 0, \quad j = 1, \ldots, k,$ por tanto $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ verifica las condiciones de Kuhn-Tucker y se convierte en el óptimo buscado (recordemos que al ser la función objetivo estrictamente convexa, la solución del programa matemático planteado es única).

5.1.2. Experimento no censurado

Cuando consideramos el experimento no censurado se sabe de antemano que la distribución que maximiza cualquier entropía es la distribución uniforme $(1/k, \ldots, 1/k)$, ahora bien, podemos construir una matriz A de tal forma que el experimento no censurado (con $c = (0, 0, \ldots, 0, 1)$), constituya un caso particular de los estudiados. Sea A la matriz:

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

Utilizando las fórmulas anteriores, resulta

$$\begin{aligned} A_{k-1} &= \frac{0^2 + 1^2}{(0+1)^2} = 1 \\ A_j &= \frac{(1 + (k-j-1)) \cdot 0 + 1 \cdot 1^2}{1^2} = 1, \qquad j = 1, \ldots, k-2 \end{aligned}$$

por tanto

$$\begin{aligned} \theta_k^* &= \frac{1}{A_1 + \cdots + A_{k-1} + 1} = \frac{1}{k} \\ \theta_{k-1}^* &= \frac{A_{k-1}}{A_1 + \cdots + A_{k-1} + 1} = \frac{1}{k} \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \theta_1^* &= \frac{A_1}{A_1 + \cdots + A_{k-1} + 1} = \frac{1}{k} \end{aligned}$$

5.1.3. Casos particulares

En esta sección se consideran tres distribuciones de censura ordenadas estocásticamente[1] $c^{(1)} \succeq c^{(2)} \succeq c^{(3)}$.

a) $c^{(1)} = \left(\frac{1}{2(k-1)}, \frac{1}{2(k-1)}, \ldots, \frac{1}{2(k-1)}, \frac{1}{2} \right), \qquad k = 2, 3, \ldots$

La matriz A que se forma es

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{k-2}{2(k-1)} + \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{k-3}{2(k-1)} + \frac{1}{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{2(k-1)} + \frac{1}{2} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} \\ 0 & \frac{1}{2(k-1)} & \frac{1}{2(k-1)} & \cdots & \frac{1}{2(k-1)} & \frac{1}{2(k-1)} \\ 0 & 0 & \frac{1}{2(k-1)} & \cdots & \frac{1}{2(k-1)} & \frac{1}{2(k-1)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{2(k-1)} & \frac{1}{2(k-1)} \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2(k-1)} \end{pmatrix}$$

en este caso se obtiene:

$$A_{k-1} = \frac{(k-1)^2 + 1}{k^2}$$

$$A_j = \frac{\left(1 + \sum_{l=j+1}^{k-1} A_l \right) + A_{j+1} \cdot (2k - j - 2)^2}{(2k - j - 1)^2}, \qquad j = 1, \ldots, k-2$$

[1]Para medidas de información paramétricas con la propiedad de suficiencia de experimentos la información acerca del parámetro θ aumenta cuando la censura aumenta estocásticamente

Ejemplo para $k = 4$

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{5}{6} & 0 & 0 \\ 0 & 0 & \frac{4}{6} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{1}{6} \end{pmatrix}$$

en este caso

$$\begin{aligned} A_3 &= \frac{(4-1)^2+1}{16} = 0.625 \\ A_2 &= \frac{(1+0.625)+0.625 \cdot 16}{25} = 0.465 \\ A_1 &= \frac{(1+0.625+0.465)+0.465 \cdot 25}{36} = 0.381 \end{aligned}$$

por tanto

$$\begin{aligned} \theta_4^* &= \frac{1}{A_1+A_2+A_3+1} = \frac{1}{0.381+0.465+0.625+1} = 0.405 \\ \theta_3^* &= \frac{A_3}{A_1+A_2+A_3+1} = \frac{0.625}{0.381+0.465+0.625+1} = 0.253 \\ \theta_2^* &= \frac{A_2}{A_1+A_2+A_3+1} = \frac{0.465}{0.381+0.465+0.625+1} = 0.188 \\ \theta_1^* &= \frac{A_1}{A_1+A_2+A_3+1} = \frac{0.381}{0.381+0.465+0.625+1} = 0.154 \end{aligned}$$

$$\boxed{\theta^* = (0.154,\ 0.188,\ 0.253,\ 0.405)}$$

$$\omega^* = A\theta = (0.154,\ 0.156,\ 0.169,\ 0.202,\ 0.141,\ 0.11,\ 0.068)$$

Se pueden utilizar los resultados del capítulo III teniendo en cuenta que

$$2A'A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1.4444 & 0.0556 & 0.0556 \\ 0 & 0.0556 & 1 & 0.1112 \\ 0 & 0.0556 & 0.1112 & 0.6666 \end{pmatrix}$$

luego

$$\theta_1^* = \frac{\begin{vmatrix} 2/7 & 0 & 0 & 0 & 1 \\ 2/7 & 1.4444 & 0.0556 & 0.0556 & 1 \\ 2/7 & 0.0556 & 1 & 0.1112 & 1 \\ 2/7 & 0.0556 & 0.1112 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.4444 & 0.0556 & 0.0556 & 1 \\ 0 & 0.0556 & 1 & 0.1112 & 1 \\ 0 & 0.0556 & 0.1112 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.154; \qquad \theta_2^* = \frac{\begin{vmatrix} 2 & 2/7 & 0 & 0 & 1 \\ 0 & 2/7 & 0.0556 & 0.0556 & 1 \\ 0 & 2/7 & 1 & 0.1112 & 1 \\ 0 & 2/7 & 0.1112 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.4444 & 0.0556 & 0.0556 & 1 \\ 0 & 0.0556 & 1 & 0.1112 & 1 \\ 0 & 0.0556 & 0.1112 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.188$$

$$\theta_3^* = \frac{\begin{vmatrix} 2 & 0 & 2/7 & 0 & 1 \\ 0 & 1.4444 & 2/7 & 0.0556 & 1 \\ 0 & 0.0556 & 2/7 & 0.1112 & 1 \\ 0 & 0.0556 & 2/7 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.4444 & 0.0556 & 0.0556 & 1 \\ 0 & 0.0556 & 1 & 0.1112 & 1 \\ 0 & 0.0556 & 0.1112 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.253; \qquad \theta_4^* = \frac{\begin{vmatrix} 2 & 0 & 0 & 2/7 & 1 \\ 0 & 1.4444 & 0.0556 & 2/7 & 1 \\ 0 & 0.0556 & 1 & 2/7 & 1 \\ 0 & 0.0556 & 0.1112 & 2/7 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.4444 & 0.0556 & 0.0556 & 1 \\ 0 & 0.0556 & 1 & 0.1112 & 1 \\ 0 & 0.0556 & 0.1112 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.405$$

$$\boxed{\theta^* = (0.154,\ 0.188,\ 0.253,\ 0.405)}$$

$$\omega^* = A\theta = (0.154,\ 0.156,\ 0.169,\ 0.202,\ 0.141,\ 0.11,\ 0.068)$$

b) Distribución uniforme para c :

$$c^{(2)} = \left(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k}\right)$$

La matriz A que se obtiene es

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{k-1}{k} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{k-2}{k} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{k} \\ 0 & \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} & \frac{1}{k} \\ 0 & 0 & \frac{1}{k} & \cdots & \frac{1}{k} & \frac{1}{k} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{k} & \frac{1}{k} \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{k} \end{pmatrix}$$

en este caso

$$A_{k-1} = \frac{1}{2} \qquad \text{(independientemente del valor de } k\text{)}$$

$$A_j = \frac{\left(1 + \sum_{l=j+1}^{k-1} A_l\right) + A_{j+1} \cdot (k-j)^2}{(k-j+1)^2}, \qquad j = 1, \ldots, k-2$$

Ejemplo para $k = 4$

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{2}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

en este caso

$$\begin{aligned} A_3 &= 0.5 \\ A_2 &= \frac{(1+0.5)+0.5\cdot 4}{9} = 0.3889 \\ A_1 &= \frac{(1+0.5+0.3889)+0.3889\cdot 9}{16} = 0.3368 \end{aligned}$$

por tanto

$$\begin{aligned} \theta_4^* &= \frac{1}{A_1+A_2+A_3+1} = \frac{1}{0.3368+0.3889+0.5+1} = 0.45 \\ \theta_3^* &= \frac{A_3}{A_1+A_2+A_3+1} = \frac{0.5}{0.3368+0.3889+0.5+1} = 0.224 \\ \theta_2^* &= \frac{A_2}{A_1+A_2+A_3+1} = \frac{0.3889}{0.3368+0.3889+0.5+1} = 0.175 \\ \theta_1^* &= \frac{A_1}{A_1+A_2+A_3+1} = \frac{0.3368}{0.3368+0.3889+0.5+1} = 0.151 \end{aligned}$$

$$\boxed{\theta^* = (0.151,\ 0.175,\ 0.224,\ 0.45)}$$

$$\omega^* = A\theta = (0.151,\ 0.1312,\ 0.112,\ 0.1125,\ 0.2123,\ 0.1685,\ 0.1125)$$

Se pueden utilizar los resultados del capítulo III teniendo en cuenta que

$$2A'A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1.25 & 0.125 & 0.125 \\ 0 & 0.125 & 0.75 & 0.25 \\ 0 & 0.125 & 0.25 & 0.5 \end{pmatrix}$$

luego

$$\theta_1^* = \frac{\begin{vmatrix} 2/7 & 0 & 0 & 0 & 1 \\ 2/7 & 1.25 & 0.125 & 0.125 & 1 \\ 2/7 & 0.125 & 0.75 & 0.25 & 1 \\ 2/7 & 0.125 & 0.25 & 0.5 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.25 & 0.125 & 0.125 & 1 \\ 0 & 0.125 & 0.75 & 0.25 & 1 \\ 0 & 0.125 & 0.25 & 0.5 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.151; \qquad \theta_2^* = \frac{\begin{vmatrix} 2 & 2/7 & 0 & 0 & 1 \\ 0 & 2/7 & 0.125 & 0.125 & 1 \\ 0 & 2/7 & 0.75 & 0.25 & 1 \\ 0 & 2/7 & 0.25 & 0.5 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.25 & 0.125 & 0.125 & 1 \\ 0 & 0.125 & 0.75 & 0.25 & 1 \\ 0 & 0.125 & 0.25 & 0.5 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.175$$

$$\theta_3^* = \frac{\begin{vmatrix} 2 & 0 & 2/7 & 0 & 1 \\ 0 & 1.25 & 2/7 & 0.125 & 1 \\ 0 & 0.125 & 2/7 & 0.25 & 1 \\ 0 & 0.125 & 2/7 & 0.5 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.25 & 0.125 & 0.125 & 1 \\ 0 & 0.125 & 0.75 & 0.25 & 1 \\ 0 & 0.125 & 0.25 & 0.5 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.224; \qquad \theta_4^* = \frac{\begin{vmatrix} 2 & 0 & 0 & 2/7 & 1 \\ 0 & 1.25 & 0.125 & 2/7 & 1 \\ 0 & 0.125 & 0.75 & 2/7 & 1 \\ 0 & 0.125 & 0.25 & 2/7 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.25 & 0.125 & 0.125 & 1 \\ 0 & 0.125 & 0.75 & 0.25 & 1 \\ 0 & 0.125 & 0.25 & 0.5 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.45$$

$$\boxed{\theta^* = (0.151,\ 0.175,\ 0.224,\ 0.45)}$$

$$\omega^* = A\theta = (0.151,\ 0.1312,\ 0.112,\ 0.1125,\ 0.2123,\ 0.1685,\ 0.1125)$$

c) $c^{(3)} = \left(\frac{1}{k-1}\,,\, \frac{1}{k-1}\,,\cdots,\, \frac{1}{k-1}\,,\, 0 \right), \qquad k = 2, 3, \ldots$

La matriz A que se obtiene es

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{k-2}{k-1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{k-3}{k-1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{k-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{k-1} & \frac{1}{k-1} & \cdots & \frac{1}{k-1} & \frac{1}{k-1} \\ 0 & 0 & \frac{1}{k-1} & \cdots & \frac{1}{k-1} & \frac{1}{k-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{k-1} & \frac{1}{k-1} \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{k-1} \end{pmatrix}$$

en este caso $\omega = (\omega_1, \ldots, \omega_k, \omega_{k+1}, \ldots, \omega_{2k-1})$, $\omega_k = \theta_k c_k = 0, \ \forall \theta$.

$$A_{k-1} = 1 \qquad \text{(independientemente del valor de } k)$$

$$A_j = \frac{\left(1 + \sum_{l=j+1}^{k-1} A_l\right) + A_{j+1} \cdot (k-j-1)^2}{(k-j)^2}\,, \qquad j = 1, \ldots, k-2$$

Ejemplo para $k = 4$

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} \end{pmatrix}$$

en este caso

$$\begin{aligned} A_3 &= 1 \\ A_2 &= \frac{(1+1)+1\cdot 1}{4} = 0.75 \\ A_1 &= \frac{(1+1+0.75)+0.75\cdot 4}{9} = 0.6389 \end{aligned}$$

por tanto

$$\begin{aligned} \theta_4^* &= \frac{1}{A_1+A_2+A_3+1} = \frac{1}{0.6389+0.75+1+1} = 0.295 \\ \theta_3^* &= \frac{A_3}{A_1+A_2+A_3+1} = \frac{1}{0.6389+0.75+1+1} = 0.295 \\ \theta_2^* &= \frac{A_2}{A_1+A_2+A_3+1} = \frac{0.75}{0.6389+0.75+1+1} = 0.221 \\ \theta_1^* &= \frac{A_1}{A_1+A_2+A_3+1} = \frac{0.6389}{0.6389+0.75+1+1} = 0.189 \end{aligned}$$

$$\boxed{\theta^* = (0.189,\ 0.221,\ 0.295,\ 0.295)}$$

$$\omega^* = A\theta = (0.189,\ 0.147,\ 0.098,\ 0,\ 0.27,\ 0.198,\ 0.098)$$

Se pueden utilizar los resultados del capítulo III teniendo en cuenta que

$$2A'A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1.1112 & 0.2222 & 0.2222 \\ 0 & 0.2222 & 0.6666 & 0.4444 \\ 0 & 0.2222 & 0.4444 & 0.6666 \end{pmatrix}$$

luego

$$\theta_1^* = \frac{\begin{vmatrix} 2/7 & 0 & 0 & 0 & 1 \\ 2/7 & 1.1112 & 0.2222 & 0.2222 & 1 \\ 2/7 & 0.2222 & 0.6666 & 0.4444 & 1 \\ 2/7 & 0.2222 & 0.4444 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.1112 & 0.2222 & 0.2222 & 1 \\ 0 & 0.2222 & 0.6666 & 0.4444 & 1 \\ 0 & 0.2222 & 0.4444 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.189; \qquad \theta_2^* = \frac{\begin{vmatrix} 2 & 2/7 & 0 & 0 & 1 \\ 0 & 2/7 & 0.2222 & 0.2222 & 1 \\ 0 & 2/7 & 0.6666 & 0.4444 & 1 \\ 0 & 2/7 & 0.4444 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.1112 & 0.2222 & 0.2222 & 1 \\ 0 & 0.2222 & 0.6666 & 0.4444 & 1 \\ 0 & 0.2222 & 0.4444 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.221$$

$$\theta_3^* = \frac{\begin{vmatrix} 2 & 0 & 2/7 & 0 & 1 \\ 0 & 1.1112 & 2/7 & 0.2222 & 1 \\ 0 & 0.2222 & 2/7 & 0.4444 & 1 \\ 0 & 0.2222 & 2/7 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.1112 & 0.2222 & 0.2222 & 1 \\ 0 & 0.2222 & 0.6666 & 0.4444 & 1 \\ 0 & 0.2222 & 0.4444 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.295; \qquad \theta_4^* = \frac{\begin{vmatrix} 2 & 0 & 0 & 2/7 & 1 \\ 0 & 1.1112 & 0.2222 & 2/7 & 1 \\ 0 & 0.2222 & 0.6666 & 2/7 & 1 \\ 0 & 0.2222 & 0.4444 & 2/7 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 2 & 0 & 0 & 0 & 1 \\ 0 & 1.1112 & 0.2222 & 0.2222 & 1 \\ 0 & 0.2222 & 0.6666 & 0.4444 & 1 \\ 0 & 0.2222 & 0.4444 & 0.6666 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}} = 0.295$$

$$\boxed{\theta^* = (0.189,\ 0.221,\ 0.295,\ 0.295)}$$

$$\omega^* = A\theta = (0.189,\ 0.147,\ 0.098,\ 0,\ 0.27,\ 0.198,\ 0.098)$$

5.2. Formulación del programa [I] para la entropía de Shannon

Dadas las caraterísticas que presenta en su formulación la entropía de Shannon, es posible obtener la solución del programa

$$\left.\begin{array}{l} \text{máx } H_{Sh}(A\theta) \\ s.a. \\ \theta_j \geq 0 \quad j = 1, \ldots, k \\ \sum_{j=1}^{k} \theta_j = 1 \end{array}\right\} \qquad \text{[I]}$$

(con H_{Sh} entropía de Shannon con logaritmos naturales) actuando de forma similar a como se ha hecho en el apartado anterior.

La función objetivo es cóncava, diferenciable como función de θ y el conjunto de soluciones factibles $\triangle_k$ es cerrado, acotado y convexo luego se trata, de un programa convexo para máximo. La función objetivo es estrictamente cóncava pues la matriz A es de rango completo como ya se ha visto anteriormente.

5.2.1. Resolución del programa

Por el teorema de Weierstrass sabemos que existe solución y además sabemos que va a ser única, pues la función es estrictamente cóncava (en cualquier situación). Para encontrar la solución, utilizamos las técnicas de programación matemática descritas en los capítulos anteriores, en concreto el Teorema de Kuhn-Tucker. Las condiciones necesarias y suficientes de Kuhn-Tucker que debe verificar un punto candidato a óptimo (en este caso global) son las siguientes:

$$(1) \quad \frac{\partial L}{\partial \theta_j}(\theta^*) = 0 \qquad j = 1, \ldots, k$$

$$(2) \quad \lambda_j \, \theta_j^* = 0 \qquad j = 1, \ldots, k$$

$$(3) \quad \theta_j^* \geq 0 \qquad j = 1, \ldots, k$$

$$(4) \quad \lambda_j \geq 0 \qquad j = 1, \ldots, k$$

$$(5) \quad \sum_{j=1}^{k} \theta_j^* = 1$$

siendo L la función Lagrangiana:

$$\begin{aligned} L &= -\theta_1 log\theta_1 - \theta_{(2)} c_1 log(\theta_{(2)} c_1) \cdots - \theta_{k-1} c_{(k-1)} log(\theta_{k-1} c_{(k-1)}) \\ &\quad -\theta_k c_{k-1} log(\theta_k c_{k-1}) - \theta_k c_k log(\theta_k c_k) + \sum_{j=1}^{k} \lambda_j \theta_j + \mu \left(\sum_{j=1}^{k} \theta_j - 1 \right) \end{aligned}$$

Para resolver analíticamente el programa, hay que encontrar las soluciones del conjunto de ecuaciones formado por las condiciones de Kuhn-Tucker (1) a (5) anteriores, es decir, hay un total de $4k + 1$ condiciones.

Las hipótesis que podemos hacer sobre los valores que toman los λ_j, $j = 1, \ldots, k$ y que dan lugar a los casos $1, \ldots, k$, ya han sido analizadas detalladamente en el capítulo III. Seguidamente demostraremos que la solución óptima que buscamos se encuentra siempre en el caso 1.

Caso 1. $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$. La función Lagrangiana que queda es:

$$\begin{aligned} L &= -\theta_1 log\theta_1 - \theta_{(2)} c_1 log(\theta_{(2)} c_1) \cdots - \theta_{k-1} c_{(k-1)} log(\theta_{k-1} c_{(k-1)}) \\ &\quad -\theta_k c_{k-1} log(\theta_k c_{k-1}) - \theta_k c_k log(\theta_k c_k) + \mu \left(\sum_{j=1}^{k} \theta_j - 1 \right) \end{aligned}$$

Las derivadas parciales de la función L son

$$\frac{\partial L}{\partial \theta_1} = -log\theta_1 - 1 + \mu$$

$$\frac{\partial L}{\partial \theta_2} = -c_1 log(\theta_{(2)} c_1) - c_1 - c_{(2)} log(\theta_2 c_{(2)}) - c_{(2)} + \lambda$$

para $2 < j \leq k-1$

$$\frac{\partial L}{\partial \theta_j} = c_1 log(\theta_{(2)} c_1) - c_1 \ \cdots \ - c_{j-1} log(\theta_{(j)} c_{j-1}) - c_{j-1} - c_{(j)} log(\theta_j c_{(j)}) - c_{(j)} + \mu$$

y cuando $j = k$

$$\begin{aligned} \frac{\partial L}{\partial \theta_k} = & -c_1 log(\theta_{(2)} c_1) - c_1 \ \cdots \ - c_{k-2} \log(\theta_{(k-1)} c_{k-2}) - c_{k-2} - c_{k-1} log(\theta_k c_{k-1}) - c_{k-1} \\ & -c_k log(\theta_k c_k) - c_k + \mu \end{aligned}$$

que al igualarlas a 0 forman el sistema:

$$\left.\begin{array}{rcl} -log\theta_1 - 1 + \mu & = & 0 \\ -c_1 log\theta_{(2)} c_1 - c_1 - c_{(2)} log(\theta_2 c_{(2)}) - c_{(2)} + \mu & = & 0 \\ \vdots \qquad\qquad\qquad \vdots & & \vdots \\ c_1 log(\theta_{(2)} c_1) - c_1 \cdots - c_{j-1} log(\theta_{(j)} c_{j-1}) - c_{j-1} - c_{(j)} log(\theta_j c_{(j)}) - c_{(j)} + \mu & = & 0 \\ \vdots \qquad\qquad\qquad \vdots & & \vdots \\ -c_1 log(\theta_{(2)} c_1) - c_1 \ \cdots \ - c_{k-2} \log(\theta_{(k-1)} c_{k-2}) - c_{k-2} - c_{k-1} log(\theta_k c_{k-1}) - c_{k-1} - c_k log(\theta_k c_k) - c_k + \mu & = & 0 \end{array}\right\}$$

Para resolverlo, procedemos de la misma forma que en el apartado anterior igualando de dos en dos las ecuaciones que forman el sitema comenzando por las dos últimas

$\frac{\partial L}{\partial \theta_{k-1}} = \frac{\partial L}{\partial \theta_k}$ simplificando queda:

$$
\begin{aligned}
-c_{(k-1)}log(\theta_{k-1}c_{(k-1)}) &= -c_{k-1}log(\theta_k c_{k-1}) - c_k log(\theta_k c_k) \\
-c_{(k-1)}logc_{(k-1)} - c_{(k-1)}log\theta_{k-1} &= -c_{k-1}logc_{k-1} - c_{k-1}log\theta_k - c_k logc_k - c_k log\theta_k \\
-c_{(k-1)}log\theta_{k-1} &= -c_{(k-1)}log\theta_k - c_{k-1}logc_{k-1} - c_k logc_k + c_{(k-1)}logc_{(k-1)} \\
-c_{(k-1)}log\theta_{k-1} &= -c_{(k-1)}log\theta_k + c_{(k-1)}H\left(\frac{c_{k-1}}{c_{(k-1)}}, \frac{c_k}{c_{(k-1)}}\right) \\
c_{(k-1)}log\theta_{k-1} &= c_{(k-1)}log\theta_k - c_{(k-1)}H\left(\frac{c_{k-1}}{c_{(k-1)}}, \frac{c_k}{c_{(k-1)}}\right)
\end{aligned}
$$

si $H\left(\frac{c_{k-1}}{c_{(k-1)}}, \frac{c_k}{c_{(k-1)}}\right) = B_{k-1} \Rightarrow log\theta_{k-1} = log\theta_k + log\, e^{B_{k-1}} \Rightarrow \theta_{k-1} = \theta_k e^{B_{k-1}}$

$$\theta_k = \theta_{k-1}e^{-B_{k-1}}$$

Igualando $\frac{\partial L}{\partial \theta_{j+1}} = \frac{\partial L}{\partial \theta_j}$, $j = 1, \ldots, k-2$ queda al simplificar

$$
\begin{aligned}
-c_{(j)}log(\theta_j c_{(j)}) &= -c_j log(\theta_{(j+1)}c_j) - c_{(j+1)}log(\theta_{j+1}c_{(j+1)}) \\
-c_{(j)}logc_{(j)} - c_{(j)}log\theta_j &= -c_j logc_j - c_j log\theta_{(j+1)} - c_{(j+1)}logc_{(j+1)} - c_{(j+1)}log\theta_{j+1} \\
&= -c_j logc_j - c_j log[\theta_{j+1} + \cdots + \theta_{k+1}] - c_{(j+1)}logc_{(j+1)} - c_{(j+1)}log\theta_{j+1} \\
&= -c_j logc_j - c_j log[\theta_{j+1}(1 + e^{-B_{j+1}} + e^{-B_{j+2}} + \cdots + e^{-B_{k-1}})] - c_{(j+1)}logc_{(j+1)} - c_{(j+1)}log\theta_{j+1} \\
-c_{(j)}log\theta_j &= -c_{(j)}log\theta_{j+1} - c_j log(1 + e^{-B_{j+1}} + e^{-B_{j+2}} + \cdots + e^{-B_{k-1}}) - c_{(j+1)}logc_{(j+1)} \\
&\quad -c_{(j+1)}log\theta_{j+1} + c_{(j)}logc_{(j)} \\
-c_{(j)}log\theta_j &= -c_{(j)}log\theta_{j+1} - c_j log(1 + e^{-B_{j+1}} + e^{-B_{j+2}} + \cdots + e^{-B_{k-1}}) + c_{(j)}H\left(\frac{c_j}{c_{(j)}}, \frac{c_{(j+1)}}{c_{(j)}}\right) \\
log\theta_j &= log\theta_{j+1} + \frac{c_j}{c_{(j)}}log(1 + e^{-B_{j+1}} + e^{-B_{j+2}} + \cdots + e^{-B_{k-1}}) - H\left(\frac{c_j}{c_{(j)}}, \frac{c_{(j+1)}}{c_{(j)}}\right)
\end{aligned}
$$

si llamamos $B_j = \frac{c_j}{c_{(j)}}log(1 + e^{-B_{j+1}} + e^{-B_{j+2}} + \cdots + e^{-B_{k-1}}) - H\left(\frac{c_j}{c_{(j)}}, \frac{c_{(j+1)}}{c_{(j)}}\right)$ queda

$$\boxed{\theta_{j+1} = \theta_j e^{-B_j}, \quad j = 1, \ldots, k-1}$$

con

$$\boxed{B_j = \frac{c_j}{c_{(j)}}\, log\left(1 + \sum_{l=j+1}^{k-1} e^{-B_l}\right) - H\left(\frac{c_j}{c_{(j)}},\ \frac{c_{(j+1)}}{c_{(j)}}\right) \quad j = 1, \ldots, k-1}$$

De $\theta_1 + \cdots + \theta_k = 1$ se obtiene sustituyendo:

$$\theta_1 + \theta_1 e^{-B_1} + \theta_1 e^{-(B_1+B_2)} + \cdots + \theta_1 e^{-(B_1+B_2+\cdots+B_{k-1})} = 1$$

por tanto

$$\begin{aligned}
\theta_1^* &= \frac{1}{1 + e^{-B_1} + e^{-(B_1+B_2)} + \cdots + e^{-(B_1+B_2+\cdots+B_k)}} \\
\theta_2^* &= \frac{e^{-B_1}}{1 + e^{-B_1} + e^{-(B_1+B_2)} + \cdots + e^{-(B_1+B_2+\cdots+B_k)}} \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
\theta_k^* &= \frac{e^{-(B_1+B_2+\cdots+B_k)}}{1 + e^{-B_1} + e^{-(B_1+B_2)} + \cdots + e^{-(B_1+B_2+\cdots+B_k)}}
\end{aligned}$$

como se observa, $\theta_j^* > 0, \;\; j = 1, \ldots, k,$ luego $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ verifica las condiciones de Kuhn- Tucker y se convierte en el óptimo buscado.

5.3. Resumen

En las tablas siguientes se muestran los valores de la entropía de Shannon, y de la entropía de Havrda y Charvát de grado 2, para las distribuciones de censura vistas anteriormente y varios θ incluyendo el modelo no censurado $c = (0, 0, \ldots, 0, 1)$; $\theta^{(*)}$ solución aproximada y θ^*_{Sh} solución del programa [I] que maximiza dichas entropías. Todos los resultados se refieren a $k = 4$.

Tabla 1. Entropías de Shannon (logarítmo natural).

$\mathrm{H}(\omega(\theta))$	$\theta^{(1)}$	$\theta^{(2)}$	$\theta^{(3)}$	$\theta^{(*)}$	θ^*_{Sh}
c	0.940	0.826	1.279	1.3863	1.3863
$c^{(1)}$	1.118	1.734	1.888	1.9033	1.9081
$c^{(2)}$	1.122	1.792	1.900	1.9177	1.9179
$c^{(3)}$	1.068	1.511	1.698	1.7261	1.7269

$$
\begin{aligned}
c &= (0,\ 0,\ 0,\ 1) & \theta^{(1)} &= (7/10,\ 1/10,\ 1/10,\ 1/10)\\
c^{(1)} &= (1/6,\ 1/6,\ 1/6,\ 1/2) & \theta^{(2)} &= (1/20,\ 2/20,\ 2/20,\ 15/20)\\
c^{(2)} &= (1/4,\ 1/4,\ 1/4,\ 1/4) & \theta^{(3)} &= (1/10,\ 2/10,\ 3/10,\ 4/10)\\
c^{(3)} &= (1/3,\ 1/3,\ 1/3,\ 0) & \theta^{(*)} &= (\theta^*_1,\ \theta^*_2,\ \theta^*_3,\ \theta^*_4)
\end{aligned}
$$

Para $c^{(1)}$

$$
\begin{aligned}
\theta^*_H &= (0.1491, 0{,}1807, 0.2433, 0.4269)\\
\omega^*_H = A\theta^*_H &= (0.1490, 0.1506, 0.1622, 0.2134, 0.1418, 0.1117, 0.0713)
\end{aligned}
$$

Para $c^{(2)}$

$$
\begin{aligned}
\theta^*_H &= (0.1469, 0{,}1732, 0.2266, 0.4533)\\
\omega^*_H = A\theta^*_H &= (0.1469, 0.1297, 0.1134, 0.1134, 0.2132, 0.1701, 0.1133)
\end{aligned}
$$

Para $c^{(3)}$

$$
\begin{aligned}
\theta^*_H &= (0.1735, 0{,}2067, 0.3099, 0.3099)\\
\omega^*_H = A\theta^*_H &= (0.1735, 0.1378, 0.1033, 0, 0.2755, 0.2066, 0.1033)
\end{aligned}
$$

Tabla 2. Entropías de Havrda y Charvát ($s = 2$).

$\mathrm{H}(\omega(\theta))$	$\theta^{(1)}$	$\theta^{(2)}$	$\theta^{(3)}$	$\theta^{(*)} = \theta^*_H$
c	0.960	0.830	1.400	1.500
$c^{(1)}$	0.984	1.570	1.683	1.691
$c^{(2)}$	0.985	1.635	1.687	1.697
$c^{(3)}$	0.978	1.497	1.600	1.623

$$
\begin{aligned}
c &= (0,\, 0,\, 0,\, 1) & \theta^{(1)} &= (7/10,\, 1/10,\, 1/10,\, 1/10)\\
c^{(1)} &= (1/6,\, 1/6,\, 1/6,\, 1/2) & \theta^{(2)} &= (1/20,\, 2/20,\, 2/20,\, 15/20)\\
c^{(2)} &= (1/4,\, 1/4,\, 1/4,\, 1/4) & \theta^{(3)} &= (1/10,\, 2/10,\, 3/10,\, 4/10)\\
c^{(3)} &= (1/3,\, 1/3,\, 1/3,\, 0) & \theta^{(*)} &= (\theta^*_1,\, \theta^*_2,\, \theta^*_3,\, \theta^*_4)
\end{aligned}
$$

Apéndice A

Matrices y Formas cuadráticas

A.1. Matrices

Denotaremos por $\mathcal{M}_{m\times n}$ el conjunto de todas las matrices de orden $m \times n$ y por $\mathcal{M}_n$ el conjunto de todas las matrices cuadradas con n filas.

Definición A.1.1. *Rango de una Matriz*

Dada una matriz cualquiera A de orden $m \times n$ se denomina rango de la matriz A y se nota por $rg(A)$ al máximo número de vectores, ya sean filas o columnas de A linealmente independientes, pues este número coincide en ambos casos.

Dada una matriz $A \in \mathcal{M}_{m\times n}$ atendiendo a su rango se pueden distinguir los siguientes tipos de matrices:

Si $m \neq n$ y $rg(A) = \text{mín}\{m, n\}$ se dice que A es de *rango completo.*

Si $m = n$ y $rg(A) = n$ se dice que A es *no singular o regular.*

Si $m = n$ y $rg(A) < n$ se dirá que A es *singular.*

Proposición A.1.1.

Dada $A \in \mathcal{M}_{m\times n}$ se verifica que $rg(A) = rg(A'A) = rg(AA')$. En particular, si $m > n$ y $rg(A) = n$, la matriz $A'A$ es no singular. Se puede ver la demostración en Barbolla y Sanz (1998).

Proposición A.1.2.

Si A es una matriz cuadrada no singular con elementos reales, tal que para cada una de sus columnas se verifica que la suma de sus elementos es 1 entonces para cada columna de A^{-1} se verifica que la suma de sus elementos es 1.

Demostración

Sea $A^{-1} = \begin{pmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & & \vdots \\ b_{k1} & \cdots & b_{kk} \end{pmatrix}$ se verifica que

$$\begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & & a_{kk} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & & \vdots \\ b_{k1} & \cdots & b_{kk} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$$

Formemos los productos que conducen a la primera columna de la matriz identidad $I_{k\times k}$

$$\begin{array}{rcl} a_{11}b_{11} + \cdots + a_{1k}b_{k1} & = & 1 \\ a_{21}b_{11} + \cdots + a_{2k}b_{k1} & = & 0 \\ \vdots \qquad\qquad\quad \vdots & & \vdots \\ a_{k1}b_{11} + \cdots + a_{kk}b_{k1} & = & 0 \\ \hline \\ b_{11}\sum_{i=1}^{k} a_{i1} + \cdots + b_{k1}\sum_{i=1}^{k} a_{ik} & = & 1 \end{array}$$

y como cada una de las columnas de A suman 1, se cumple que $b_{11} + \cdots + b_{k1} = 1$ de igual forma se demuestra para las restantes columnas de A^{-1}.

Definición A.1.2. *Producto Kronecker*

Dadas las matrices $A_{m\times n}$, $B_{p\times q}$, se define el *producto de Kronecker* de A por B, que se denota por $A \otimes B$, como la matriz de orden $mp \times nq$ dada por

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

A.2. Formas cuadráticas

Definición A.2.1. *Polinomio cuadrático*

Se dice que un *polinomio* p en las variables $x_1, x_2, \ldots, x_n$ es *cuadrático,* si cada uno de sus términos tiene grado dos, es decir

$$p_2(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

siendo los coeficientes $a_{ij} \in \mathbb{R},\ i, j = 1, \ldots, n$ y las variables $x_i,\ i = 1, \ldots, n$ con valores en $\mathbb{R}$.

Definición A.2.2. *Forma cuadrática*

Se denomina *forma cuadrática* q a toda aplicación de $\mathbb{R}^n$ en $\mathbb{R}$ que a cada vector $\bar{x} \in \mathbb{R}^n$ le hace corresponder el valor numérico dado por un polinomio cuadrático.

Definición A.2.3. *Matriz asociada a una forma cuadrática*

Dada una forma cuadrática q, definida de $\mathbb{R}^n$ en $\mathbb{R}$ la única matriz simétrica $Q \in \mathcal{M}_n$ para la que se verifica $q(\bar{x}) = \bar{x}'\, Q\, \bar{x}$ se dice que es la matriz asociada a la forma cuadrática q, denominándose expresión matricial de q a la dada a partir de la matriz simétrica Q.

Definición A.2.4. *Tipos de formas cuadráticas*

Sea q una forma cuadrática en las variables $(x_1, x_2, \ldots, x_n)$. Se dice que

1. q es *definida positiva* si y sólo si para todo $\bar{x} \in \mathbb{R}^n\ \ \bar{x} \neq 0$ se verifica que $q(\bar{x}) > 0$.

2. q es *definida negativa* si y sólo si para todo $\bar{x} \in \mathbb{R}^n\ \ \bar{x} \neq 0$ se verifica que $q(\bar{x}) < 0$.

3. q es *semidefinida positiva* si y sólo si para todo $\bar{x} \in \mathbb{R}^n\ \ q(\bar{x}) \geq 0$ y existe algún vector no nulo $\bar{x}^1$ tal que $q(\bar{x}^1) = 0$.

4. q es *semidefinida negativa* si y sólo si para todo $\bar{x} \in \mathbb{R}^n\ \ q(\bar{x}) \leq 0$ y existe algún vector no nulo $\bar{x}^2$ tal que $q(\bar{x}^2) = 0$.

5. q es *indefinida* si y sólo si existen $\bar{x}^0,\ \bar{x}^* \in \mathbb{R}^n$ tales que $q(\bar{x}^0) < 0$ y $q(\bar{x}^*) > 0$.

Definición A.2.5. *Menor principal*

Se denomina *menor principal* $D_i,\ i = 1, \ldots, n$ de una matriz $A = (a_{ij}),\ i, j = 1, \ldots, n$ a

$$D_i = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ii} \end{vmatrix}$$

Criterios de clasificación de formas cuadráticas

Teorema A.2.1. *Criterio de los menores*

Sea $q(\bar{x}) = \bar{x}' A \bar{x}$ una forma cuadrática en las variables $x_1, \ldots, x_n$. Entonces se verifica que:

1. q es *definida positiva* si y sólo si $D_i > 0,\ \ i = 1, \ldots, n.$
2. q es *definida negativa* si y sólo si $(-1)^i D_i > 0,\ \ i = 1, \ldots, n.$
3. q es *semidefinida positiva* si $D_i > 0,\ \ i = 1, \ldots, n-1$ y $D_n = |A| = 0.$
4. q es *semidefinida negativa* si $(-1)^i D_i > 0,\ \ i = 1, \ldots, n-1$ y $D_n = |A| = 0.$

Teorema A.2.2. *Criterio de los autovalores*

Sea $q(\bar{x}) = \bar{x}' A \bar{x}$ una forma cuadrática con matriz asociada A, cuyos autovalores son $\lambda_1, \lambda_2, \ldots, \lambda_n$. Entonces se verifica que

1. q es *definida positiva* si y sólo si $\lambda_i > 0,\ \ i = 1, \ldots, n.$
2. q es *definida negativa* si y sólo si $\lambda_i < 0,\ \ i = 1, \ldots, n.$
3. q es *semidefinida positiva* si y sólo si $\lambda_i \geq 0,\ \ i = 1, \ldots, n$ y, al menos existe i_0 tal que $\lambda_{i_0} = 0.$
4. q es *semidefinida negativa* si y sólo si $\lambda_i \leq 0,\ \ i = 1, \ldots, n$ y, al menos existe i_1 tal que $\lambda_{i_1} = 0.$
5. q es *indefinida* si y sólo si existen al menos i_2 e i_3 tales que $\lambda_{i_2} > 0$ y $\lambda_{i_3} < 0.$

Definición A.2.6. *Matrices definidas y semidefinidas*

Se dice que una matriz real y simétrica, de orden n, es definida positiva, definida negativa, semidefinida positiva o semidefinida negativa si lo es, respectivamente, la forma cuadrática $q : \mathbb{R}^n \to \mathbb{R}$ asociada a la matriz A en la base canónica.

Definición A.2.7. *Menor principal primario*

Dada una matriz A de orden n, se denomina *menor principal primario* de A de orden $p \leq n$, denotado por H_p, al valor del determinante de una submatriz de orden p de A, que se obtiene cuando en A se eliminan $n - p$ filas y columnas del mismo índice.

Proposición A.2.1.

Si $q(\bar{x}) = \bar{x}'A\bar{x}$ es una forma cuadrática definida positiva, entonces todo menor principal primario de A es positivo siendo además la submatriz asociada correspondiente de orden p definida positiva. La demostración puede verse en Muñoz, F. (1988)

Proposición A.2.2.

Dada la forma cuadrática $q(\bar{x}) = \bar{x}'A\bar{x}$ en las variables $\bar{x} = (x_1, \ldots, x_n)$ se tiene que:

1. La forma cuadrática q es definida positiva si y sólo si existe una matriz B de orden $m \times n$ con $m \geq n$ y $rg(B) = n$ tal que $A = B'B$.

2. Si $rg(A) = r < n$, la forma cuadrática q es semidefinida positiva si y sólo si existe una matriz B de orden $m \times n$ con $m \geq n$ y $rg(B) = r < n$ tal que $A = B'B$.

Se puede ver la demostración de esta proposición en Barbolla y Sanz (1998).

Definición A.2.8. *Formas cuadráticas restringidas*

Dadas las matrices $A_{n\times n}$ y $B_{m\times n}$ con $m < n$ y $rg(B) = m$, se dice que la forma cuadrática restringida

$$q(\bar{x}) = \bar{x}'A\bar{x}, \quad \text{sujeta a} \quad B\bar{x} = \bar{0} \quad \text{es:}$$

1. *Definida positiva* si y sólo si para todo $\bar{x} \in \mathbb{R}^n$, $\bar{x} \neq \bar{0}$ tal que $B\bar{x} = \bar{0}$ se verifica que $q(\bar{x}) > 0$.

2. *Definida negativa* si y sólo si para todo $\bar{x} \in \mathbb{R}^n$, $\bar{x} \neq \bar{0}$ tal que $B\bar{x} = \bar{0}$ se verifica

que $q(\bar{x}) < 0$.

3. *Semidefinida positiva* si y sólo si para todo $\bar{x} \in \mathbb{R}^n$, tal que $B\bar{x} = \bar{0}$ se verifica que $q(\bar{x}) \geq 0$, existiendo $\bar{x}^0 \neq \bar{0}$ con $B\bar{x}^0 = \bar{0}$ para el cual $q(\bar{x}^0) = 0$.

4. *Semidefinida negativa* si y sólo si para todo $\bar{x} \in \mathbb{R}^n$, tal que $B\bar{x} = \bar{0}$ se verifica que $q(\bar{x}) \leq 0$, existiendo $\bar{x}^* \neq \bar{0}$ con $B\bar{x}^* = \bar{0}$ para el cual $q(\bar{x}^*) = 0$.

5. *Indefinida* si y sólo si existen $\bar{x}^1$ y $\bar{x}^2$ no nulos tales que $B\bar{x}^1 = \bar{0}$ y $B\bar{x}^2 = \bar{0}$ para los que se verifica que $q(\bar{x}^1) > 0$ y $q(\bar{x}^2) < 0$.

• Si G es una matriz cuadrada de orden $n \times n$, denotaremos por

G_r la matriz de orden r formada por las r primeras filas y columnas de G.

• Si S es una matriz de orden $m \times n$ con $m < n$ denotaremos por

S_m la matriz de orden m obtenida a partir de las m primeras columnas de S.

$S_{m\times\underline{k}}$ la matriz de orden $m \times k$ formada por los elementos de las columnas $m+1, \ldots, m+k$ de S.

Lema A.2.1.

Dada la forma cuadrática restringida q indicada en la definición A.2.8. en las variables $x_1, x_2, \ldots, x_n$ se verifica que existe una forma cuadrática

$$q^*(\bar{y}) = \bar{y}\,' E\, \bar{y}$$

con $\bar{y} \in \mathbb{R}^{n-m}$, tal que q y q^* son ambas del mismo tipo, siendo $E = C'AC$ y

$$C = \begin{pmatrix} -B_m^{-1} B_{m\times\underline{n-m}} \\ I_{n-m} \end{pmatrix}$$

Definición A.2.9. *Matriz orlada*

Dadas las matrices A_n y $B_{m\times n}$, llamamos *matriz* A *orlada* con B a la matriz de orden $(m+n) \times (m+n)$

$$M = \left(\begin{array}{c|c} O_m & B \\ \hline B' & A \end{array}\right)$$

siendo O_m la matriz cuadrada nula de orden m.

Nota.- Algunos autores denominan matriz orlada de A con B a:

$$\tilde{M} = \left(\begin{array}{c|c} A & B' \\ \hline B & O_m \end{array}\right)$$

Lema A.2.2.

Si para cada $i = 1, \ldots, n-m$ se nota por E_i la matriz de orden i formada por las i primeras filas y columnas de la matriz E definida en el Lema A.2.1 y por B_m y M_{2m+i} lo análogo a partir de las matrices B y M indicadas en las definiciones A.2.8 y A.2.9 respectivamente, entonces se verifica que

$$|M_{2m+i}| = (-1)^m |B_m|^2 |E_i| \qquad i = 1, \ldots, n-m.$$

Se pueden ver las demostraciones de los lemas A.2.1 y A.2.2 en Barbolla y Sanz (1998). Se obtienen resultados análogos a los expuestos, si se considera la matriz $\tilde{M}$.

Apéndice B

Espacios métricos y normados

La noción de distancia como espacio o intervalo de lugar que media entre dos cosas se presenta de forma natural en la geometría euclídea al medir las longitudes de los segmentos que unen dos puntos cualesquiera del espacio. Cuando se prescinde del soporte geométrico que hace intuitiva tal noción y se consideran sus propiedades esenciales, se obtienen los axiomas que definen una métrica en un conjunto, y aparece el concepto de espacio métrico.

B.1. Espacio métrico

Definición B.1.1. *Métrica*
Dado el conjunto E no vacío, una métrica o distancia definida en E es una aplicación $E \times E \to \mathbb{R}$, en la que a cada par ordenado (x, y) de elementos de E le corresponde un número real que cumple las condiciones:

1. $d(x, y) \geq 0$, para todos $x, y \in E$.
2. $d(x, y) = 0$ si, y sólo si, $x = y$
3. $d(x, y) = d(y, x)$, para todos $x, y \in E$.
4. $d(x, z) \leq d(x, y) + d(y, z)$, para todos $x, y, z \in E$

Espacio métrico es el par $\{E, d\}$ formado por un conjunto E no vacío y una métrica definida en el mismo.

Dos espacios métricos son distintos cuando difieren en el conjunto soporte E o cuando

teniendo el mismo soporte E, difieren en las métricas.

De acuerdo con la definición de espacio métrico, estos espacios no necesitan tener ninguna clase de estructura algebraica definida en él y por otra parte, no son topológicos; sin embargo, como la métrica permite de manera muy directa definir una base de entornos, se dice que los espacios métricos son una clase especial de espacios topológicos.

Dado el espacio métrico $\{E, d\}$, se llama:

- *Bola abierta* de centro x y radio r al conjunto

$$B(x,r) = B_r(x) = \{y : \ y \in E, \ \ d(x,y) < r\}$$

- *Bola cerrada* de centro x y radio r al conjunto

$$\bar{B}(x,r) = \bar{B}_r(x) = \{y : \ y \in E, \ \ d(x,y) \leq r\}$$

Un conjunto A de un espacio métrico $\{E, d\}$ está acotado si y sólo si existe una bola que lo contiene.

B.2. Espacios normados

Muchos de los espacios métricos que se presentan en Análisis Matemático admiten una estructura previa de espacio vectorial, y en ellos la distancia aparece estrechamente ligada a la noción de norma de un vector. Tal es el caso del conjunto $\mathbb{R}^n$ cuya estructura de espacio vectorial sobre el cuerpo $\mathbb{R}$ es evidente. Conviene, pues, distinguir una clase particular de espacios métricos que son espacios vectoriales en los que para cada vector se puede definir una norma.

Definición B.2.1. *Norma*

Dado un espacio vectorial E sobre un cuerpo $\mathbb{K}$ real o complejo; una norma definida en E es una aplicación de E en $\mathbb{R}$, en la que a cada $\bar{x} \in E$ le corresponde un número real que se designa por $||\bar{x}||$, que verifica las siguientes condiciones:

1. $||\bar{x}|| \geq 0$ para todo $\bar{x} \in E$.

2. $||\bar{x}|| = 0$ equivale a $\bar{x} = 0$.

3. $||\alpha\,\bar{x}|| = |\alpha|\,||\bar{x}||$ para cada $\bar{x} \in E$ y cada $\alpha \in \mathbb{K}$.

4. $||\bar{x} + \bar{y}|| \leq ||\bar{x}|| + ||\bar{y}||$ para cada par $\bar{x}\,,\ \bar{y} \in E$
(desigualdad triangular de la norma).

Definición B.2.2. *Espacio normado*

Un espacio normado sobre $\mathbb{K}$ es un par $\{E,\ ||\cdot||\}$, donde E es un espacio vectorial sobre un cuerpo $\mathbb{K}$ y $||\cdot||$ una norma definida en E.

En particular, son espacios normados:

1. El espacio vectorial $\mathbb{R}$ con la norma del valor absoluto.
2. El espacio vectorial $\mathbb{R}^n$ con la norma euclídea: $||\bar{x}|| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$.

A todo espacio normado se le puede dotar, de manera natural, de una estructura de espacio métrico:

Proposición B.2.1.

Si $||\bar{x}||$ es una norma en un espacio vectorial E, la aplicación en la que a cada par ordenado de elementos $\bar{x},\, \bar{y} \in E$ le corresponde $||\bar{x} - \bar{y}||$ es una distancia $d(\bar{x}, \bar{y})$ definida en E.

Demostración.
$d(\bar{x}, \bar{y}) = ||\bar{x} - \bar{y}|| \geq 0$. Según 2, si $d(\bar{x}, \bar{y}) = ||\bar{x} - \bar{y}|| = 0$, es $\bar{x} - \bar{y} = 0$; según 3 es $d(\bar{x}, \bar{y}) = ||\bar{x} - \bar{y}|| = ||(-1)(\bar{y} - \bar{x})|| = |-1|\,||\bar{y} - \bar{x}|| = d(\bar{y}, \bar{x})$; y finalmente de 4 resulta la desigualdad triangular

$$d(\bar{x}, \bar{z}) = ||(\bar{x} - \bar{y}) + (\bar{y} - \bar{z})|| \leq ||\bar{x} - \bar{y}|| + ||\bar{y} - \bar{z}|| = d(\bar{x}, \bar{y}) + d(\bar{y}, \bar{z}).$$

Por tanto todo espacio normado E se considera como espacio métrico, con la distancia $d(\bar{x}, \bar{y}) = ||\bar{x} - \bar{y}||$.

REFERENCIAS

Aczél, J. D.; Daróczy, Z. : Uber verallgemeinerte quasilineare Mittelwerte die mit Gewichtsfunktionnen gebildet sind. *Publications Mathematicae* **1963**, *10*, 171-190.

Aczél, J. D.; Daróczy, Z. : *On Measures of Information and their Characterizations.* Academic Press, New York **1975**.

Arimoto, S. : Informatión theoretic considerations on estimation problems. *Information and Control*, **1971**, *19*, 181-194.

Barbolla, R.; Sanz, P. : *La Concavidad en un Modelo Económico.* Ed. Pirámide, **1995**.

Barbolla, R.; Sanz, P. : *Álgebra lineal y teoría de matrices.* Ed. Prentice Hall, **1998**.

Barbolla, R.; Cerdá, E.; Sanz, P. : *Optimización: Cuestiones, ejercicios y aplicaciones a la economía*. Prentice Hall. **2000**

Berkson, J.; Gage, R. : Calculation of survival rates for cancer. *Proceeding of Staff Meetings, of the Mayo Clinic*, **1950**, *25*, 270-286.

Belis, M.; Guiasu, S. : Quantitative-qualitative measure of information in cybernetic systems. *IEEE Transactions on Information Theory*, **1968**, *IT-14*, 593-594.

Blumer, A. C.; McEliece, R. J. : The Rényi redundancy of generalized huffman codes. *IEEE Transactions on Information Theory* **1988** *IT-34*, 1242-1249.

Boltzmann, L. : Vorlesungen uber Gastheorie. J. A. Barth. Leipzig. **1896**.

Campbell, L. L. : A coding theorem and Rényi's entropy. *Information and Control* **1965**, *23*, 423-429.

Campbell, L. L. : The relation between Information Theory and the Differential Geometry approach to Statistics. *Information Sciences*, **1985**, *35*, 199-210.

Chaundy, T. W.; McLeod, J.B. : On a functional equation. *Proceedings of Edinburgh Mathematical Society, Edinburgh Math. Notes*, **1960**, *43*, 7-8.

Cox, D. R. : Some simple approximate test for Poisson variates. *Biometrika* **1953**, *40*, 354-360.

Csiszár, I. : Information measures: A critial survey. *Trans. of the 7th Prague Conferen.* **1974**, 83-86.

Daróczy, Z. : Generalized information functions. *Information and Control*, **1970**, *16*, 299-310.

Davis, D. J. : An analysis of some failure data. *J. Am. Stat. Assoc.*, **1952**, *47*, 113-150.

Ebrahimi, N. : The maximum entropy method for lifetime distributions. *Sankhya.* **2000**, *A*, 236-243.

Emptoz, H. : Informatión de type β intégrant un concept d'utilité. *C. R. Acad. Sci. Paris Ser.* **1976** 911-914.

Epstein, B. : The exponential distribution and its role in life-testing. *Ind. Qual. Control.* **1958**, *15*, 2-7.

Esteban, M. D.; Morales, D.: A summary on entropy statistics. *Kybernetika.* **1995**, Vol. 32, N. 4, 337-350.

Feigl, P.; Zelen, M. : Estimation of exponential survival probabilities with concomitant information. *Biometrics* **1965**, *21*, 826-838.

Feinstein, F. : *Foundations of Information Theory,* McGraw-Hill, New York **1958**.

Fernández C., Hernández, F. J.,Vegas J. M. : *Cálculo diferencial de varias variables.* Ed. Thomson **2002**.

Ferreri, C. : Hypoentropy and related heterogenity divergency and information measures. *Statistica*, **1980**, *40*, 155-168.

Gehan, E. A. : A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **1965**, 52 (1 and 2), 203-223.

Gil, M. A.; Pérez, R.; Gil, P. : A family of measures of uncertainty involving utilities: Definitions, properties and statistical inferences. *Metrika*, **1989**, *36*, 129-147.

Guiasu, S. : *Informatión Theory with Applications*, McGraw-Hill, New York **1977**.

Gumbel, E. J. : *Statistics of Extremes*. New York: Columbia University Press. **1958**.

Gupta, S. S.; Groll, P. A. : Gamma distribution in acceptance sampling based on life test. *J. Am. Stat. Assoc.,* **1961**, *56*, 942-970.

Hald, A. : *A History of Probability and Statistics and their Applications before 1750.* John Wiley and Sons, Inc. New York, USA, **1990**.

Hartley, R.V.L. : Transmission of Information. *Bell System Technical Journal*, **1928**, *7*, 535-563.

Havrda, J.; Charvát, F. : Quantification method of classification processes: concept of structural a-entropy. *Kybernetika*, **1967**, *3*, 30-35.

Hosmer, D.W. Jr; Lemeshow, S. : *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley and Sons, Inc., New York, USA, **1999**.

Kalbfleisch, J. D.; Prentice, R. L. : *The statistical analysis of failure time data.* John Wiley and Sons, Inc., New York, USA, **1980**.

Kao, J. H. K. : A graphical estimation of mixed Weibull parameters in life testing of electron tubes. *Thechnometrics*, **1959**, *1*, 389-407.

Kaplan, E. L.; Meier, P. : Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **1958**, *53*, 475-481.

Kapur, J. N. : Generalized entropy of order α and type β. *Mathematical Seminar*, Delhi **1967**, *4*, 78-94.

Kapur, J. N. : Some new nonadditive measures of entropy, Bull. U.M.I. **1988**, 253-266

Kieffer, J. C. : Variable-length source coding with a cost depending only on the code-word length. *Information and Control*. **1979**, *41*, 136-146.

Lawless, J. F. : *Statistical models and methods for lifetime data.* John Wiley and Sons, Inc., New York, USA, **1982**.

Lieblein, J.; Zelen, M. : Statistical investigation of the fatigue life of deep groove ball bearings. *J. Res. Nat. Bur. Stand.*, **1956**, *57*, 273-316.

Mantel, N. : Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Rep.* **1966**, *50*(3), 163-170.

Mathai, A. M.; Rathie, P. N. : *Basic Concepts in Information Theory and Statistics.* Wiley Eastern, New Delhi **1975**.

McEliece, R. J. : *The Theory of Information and Coding. Encyclopedia of Mathematics and its Applications.* Addison-Wesley, Reading, Mass. **1977**.

Meeker, W. Q.; Nelson, W. B. : Tables for the Weibull y Smallest Extreme Value distributions. *Relia-Com Review* **1974**.

Morales, D.; Pardo, L.; Vajda, I. Uncertainty of discrete stochastic systems: General theory and statistical inference. *IEEE Transactions on System, Man and Cybernetics* **1996**.

Muñoz, F.; Devesa, J.; Mocholi, M.; Guerra, J. : *Manual de Álgebra Lineal* Ed. Ariel Economía **1988**.

Nyquist, H. : Certain factors affecting telegraph speed. *Bell System Technical Journal* **1924**, *3*, 324.

Nyquist, H. : Certain topics in telegraph transmission theory. *AIEEE Transactions* **1928**, *47*, 617.

Pardo, J. A. : Caracterización axiomática de la energía informacional útil. *Estadística Española*, **1985**, *108*, 107-116.

Pardo, J. A. : On the asymptotic distribution of useful Shannon entropy. *Metron*, **1993**, *LI(1-2)*, 119-137.

Pardo, J. A. : Some applications of the useful mutual information. *Applied Mathematics and Computation*, **1995**, *27*, 33-50.

Pardo, J. A.; Pardo, M. C. : Statistical applications of order α-β weighted information energy. *Applications of Mathematics*, **1995**, *40(3)*, 305-317.

Pardo, L. : Order-α weighted information energy. *Information Sciences,* **1986**, *40*, 155-164.

Peto, R.; Lee, P. : Weibull distributions for continuous carcinogensis experiments. *Biometrics.* **1973**, *29*, 457-470.

Picard, C. F. : Weighted probabilistic information measures. *Journal of Combinatorics, Information and System Sciences*, **1979**, *4*, 343-356.

Rathie, P. N. : On generalized entropy and coding theorem. *Journal of Applied Probability*, **1970**, *7*, 124-133.

Rényi, A. : On measures of entropy and information. *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, Univ. of California Press, Berkeley, **1961**, *1*, 547-561

Salicru, M.; Menendez, M. L.; Morales, D.; Pardo, L. : Asymptotic distribution of (h, ϕ)-entropies. *Communications in Statistics: Theory and Methods,* **1993** 22, 7, 2015-2031.

Sant'anna, A. P.; Taneja, I. J. : Trigonometric entropies, Jensen difference divergence measures an error bounds. *Information Sciences*, **1985**, *35*, 145-155.

Shannon, C. E. : A mathematical theory of communication. *Bell System Technical Journal* **1948**, *27*, 379-423.

Shannon, C. E. : Communication theory of secrecy systems. *Bell System Technical Journal.* **1949**, *28*, 656-715.

Sharma, B. D.; Mittal, D. P. : New nonadditive measures of inaccuracy. *Journal of Mathematical Sciences*, **1975**, *10*, 122-133.

Sharma, B. D.; Taneja, I. J. : Entropy of type (α, β) and other generalized additive measures in information theory. *Metrika*, **1975**, *22*, 205-215.

Sharma, B. D.; Taneja, I. J. : Three generalized additive measures of entropy. *Elec. Inform. Kybernet*, **1977**, *13*, 419-433.

Taneja, I. J. : Some contributions to information theory - I (A Survey): On measures of information. *J. Comb. In form and Syst. Sci.*, **1979**, 4, 253-74.

Taneja, I. J. : On Generalized Entropies with Applications. Chapter in: *Lectures in Appl. Math. and Inform.*, Ed. L.M. Ricciardi, Manchester University Press. **1990**, 107-169.

Tribus, M. : *Boelter Anniversary Volume*. McGraw-Hill. **1963**.

Turrero, A. : *Pérdida de información a causa de la censura.* Tesis Doctoral. Editorial de la U. C. M. Colección Tesis Doctorales 361/88. **1988**.

Turrero, A. : On the relative efficiency of grouped and censored survival data. *Biometrika* **1989**, *76*, 125-131.

Turrero, A. : Relative efficiency of a censored experiment in terms of Fisher Information. *Communications in Statistics: Theory and Methods* **1995**, *24,* 1169-1191.

Varma, R. S. : Generalizations of Rényi's entropy of order α. *Journal of Mathematical Sciences* **1966**, *1*, 34-48.

Weber, B.; Depew, D.; Dyke, C.; Salthe, S.; Schneider, E.; Ulanowicz, R.; Wicken, J.: Evolution in thermodynamic perspective: An ecological approach. *Biology and Philosophy* **1989**, *4*, 373-405.

Weibull, W. A : Statistical distribution function of wide applicability. *J. Appl. Mech.* **1951**, *18*, 293-297.

Whittemore, A.; Altschuler, B. : Lung cancer incidence in cigarette smokers: further analysis of Doll and Hill's data for British physicians. *Biometrics* **1976**, *32*, 805-816.

Wiener, N. : *Cybernetics*. The MIT Press and Wiley, New York **(1948)**.

Zellner, A.; Highfield, R. : Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *Journal of Econometrics.* **1988**, *37*, 195-209.