

# DISEÑO OPTIMO DE EXPERIMENTOS

**J. López Fidalgo**  
Departamento de Estadística.  
Universidad de Salamanca.



# Índice general

<b>1. Introducción</b>	<b>5</b>
1.1. Ejemplo . . . . .	5
1.2. Etapas . . . . .	7
1.3. Panorámica histórica . . . . .	7
1.4. Contexto . . . . .	9
<b>2. Estimadores de funcionales</b>	<b>15</b>
<b>3. Regresión no lineal</b>	<b>25</b>
<b>4. Criterios</b>	<b>29</b>
4.1. Introducción . . . . .	29
4.2. Funciones criterio . . . . .	30
4.2.1. D-optimización . . . . .	31
4.2.2. A-optimización . . . . .	32
4.2.3. G-optimización . . . . .	32
4.2.4. E-optimización . . . . .	33
4.2.5. c-optimización . . . . .	34
4.2.6. L-optimización . . . . .	35
4.2.7. $\Phi_p$ -optimización . . . . .	36
4.2.8. $L_p$ -optimización . . . . .	36
4.2.9. I-optimización . . . . .	36
4.2.10. MV-optimización . . . . .	37
4.2.11. Criterio minimax . . . . .	37

<b>5. Teorema de equivalencia</b>	<b>41</b>
<b>6. Construcción</b>	<b>45</b>
6.1. Aproximación . . . . .	45
6.2. Procesos iterativos . . . . .	46
6.2.1. Paso armónico . . . . .	46
6.2.2. Algoritmo ascendente . . . . .	46
6.2.3. Algoritmo acelerado . . . . .	47
6.2.4. $X$ finito . . . . .	48
<b>7. EJERCICIOS</b>	<b>53</b>

# Capítulo 1

## Introducción

Un experimento se lleva a cabo con el objeto de extraer conclusiones generales a partir de la observación de un número limitado de casos, es decir, a partir de una muestra. Interesa que los casos analizados proporcionen una información suficiente y representativa acerca de toda la población. A primera vista se aprecia que dicha fiabilidad crece con el aumento de casos observados, y es cierto, pero en la práctica dicho número estará limitado por factores económicos, temporales, de falta de recursos, etc. Además el tratamiento estadístico de estos datos se complica con el aumento del número de casos observados.

De lo dicho anteriormente se desprende la necesidad de optimizar los resultados finales empleando para ello las observaciones experimentales oportunas y estrictamente necesarias. El diseño óptimo de experimentos, como su propio nombre indica tratará de diseñar un experimento de forma que se alcance la inferencia estadística más precisa posible con el mínimo coste. Un estudio apropiado sobre el mejor diseño experimental mejora en gran medida la estimación en los modelos de regresión.

Antes de seguir adelante conviene precisar claramente qué entendemos por experimento científico. Se podría dar una definición de *experimento científico* como un conjunto, más o menos complejo, de actividades realizadas con el fin de alcanzar un conocimiento profundo acerca de un objeto. Se trata pues de un proceso cuyos resultados no se conocen de antemano con certeza.

El siguiente ejemplo nos mostrará la importancia de hacer un buen diseño del experimento:

### 1.1. Ejemplo ilustrativo

Se trata de obtener el peso de dos objetos. Para ello disponemos de una balanza tradicional con dos platos y la posibilidad de hacer dos pesadas. Se supone que el error de pesada o error experimental,  $\varepsilon$ , sigue una distribución  $\mathcal{N}(0, \sigma^2)$ , de modo que si  $P$  es el peso real de un objeto e  $y$  es la medida que da la balanza el modelo es:

$$y = P + \varepsilon$$

El diseño clásico de experimentación consistirá en pesar separadamente los dos objetos y obser-

var  $y_1$  e  $y_2$ , que serán las mejores estimaciones de los pesos. Se trata de las estimaciones máximo verosímiles, que son centradas y de varianza  $\sigma^2$ . Además dichas estimaciones son independientes, por serlo las pesadas. Eso significa que ambas estimaciones son incorreladas. En la Figura 1.1 se muestran las dos pesadas.

Figura 1.1: Diseño clásico

Veamos que existe un diseño experimental que proporciona una varianza menor de los estimadores. Consistirá en utilizar los dos objetos en las dos pesadas. En la primera se colocan los dos en el mismo plato y en la segunda en platos distintos, de modo que se obtienen los pesos:

$$y_3 = P_A + P_B + \varepsilon_1$$

$$y_4 = P_A - P_B + \varepsilon_2$$

Y la estimación mejor de los pesos (EMV) es:

$$\hat{P}_A = \frac{y_3 + y_4}{2}$$

$$\hat{P}_B = \frac{y_3 - y_4}{2}$$

Se trata de estimadores centrados cuya varianza es:

$$\text{var}(\hat{P}_A) = \text{var}(\hat{P}_B) = \frac{1}{4}2\sigma^2 = \frac{\sigma^2}{2}$$

Por tanto se consigue disminuir la varianza de los estimadores a la mitad haciendo un uso óptimo de la información experimental. En este caso las estimaciones de los dos parámetros son también incorreladas:

$$\text{cov}(\hat{P}_A, \hat{P}_B) = 1/4\{\text{cov}(y_3, y_3) - \text{cov}(y_3, y_4) + \text{cov}(y_3, y_4) - \text{cov}(y_4, y_4)\} = 0$$

y puesto que la distribución conjunta de las dos estimaciones se ha supuesto normal, entonces serán también independientes. En la Figura 1.2 se muestran estas dos pesadas.

Figura 1.2: Diseño mejorado

## 1.2. Etapas del diseño óptimo de un experimento

Mediante el experimento científico obtenemos unas observaciones que nos van a permitir ajustar un modelo al objeto de estudio. Dicho modelo debe contener una descripción del estado del objeto observado. Estamos interesados en el modelo de regresión, en el que se conoce de antemano el conjunto de puntos observables. Un estado se definirá como la función que asigna a cada punto de este conjunto la media de las cantidades observadas en él. Estas cantidades serán consideradas como variables aleatorias. La elección del modelo para la búsqueda del diseño óptimo es un problema abierto que no tiene una solución general. Interesa buscar un diseño que dé estimadores precisos para el modelo elegido, y que simultáneamente proporcione protección contra modelos inadecuados. Un buen diseño en un determinado aspecto puede no serlo tanto en otro. Por tanto el diseño óptimo de un experimento se desarrolla en las tres etapas siguientes:

1. Elección del modelo de regresión.
2. Elección de un criterio de optimización; esto exige confrontar la teoría con la situación real.
3. Utilización de un algoritmo para el cálculo del diseño óptimo. Eso necesitará la medición de la distancia que separa el diseño óptimo de los diseños que vamos obteniendo en cada paso.

## 1.3. Panorámica histórica del diseño óptimo de experimentos

Hagamos una pequeña introducción histórica para situar el tema. El primer trabajo en diseño óptimo de experimentos fue publicado el año 1918 por una mujer, Kirstine Smith [189]. Propuso un criterio para la regresión polinomial, que más tarde, en 1959, fue llamado G-optimización por Kiefer y Wolfowitz, [128]. Sin embargo ha sido a partir de los años cincuenta cuando se ha comenzado a trabajar en mayor medida en este tema. El punto de arranque para el desarrollo de esta teoría fue la matriz de dispersión obtenida por el método de los mínimos cuadrados, también llamada matriz de información. El diseño óptimo de experimentos se desarrolla en dos corrientes paralelas. Por una parte G. E. P. Box y sus seguidores (N. R. Draper, J. S. Hunter, Lucas, Wilson y otros) basan su trabajo en la matriz de dispersión para valorar la elección de los puntos de observación. Desde este punto de vista la generalización a funciones no polinómicas se hace problemática. Por otro lado J. Kiefer propondrá el empleo de funcionales de la matriz de dispersión como posibles criterios de optimización. Algunos de sus seguidores son Atwood, Covey-Crump, Silvey, Fedorov, Karlin, Studden, Whittle, Wynn,...

En 1943 Wald, [210], establece el criterio de maximización del determinante de la matriz de información. Más tarde Kiefer y Wolfowitz le darán el nombre de D-optimización y extenderán su utilización al modelo de regresión más general. Diez años después Chernoff, [56], utiliza el teorema de Taylor para linealizar modelos no lineales. Emplea para ello un valor inicial de los parámetros y el criterio de la maximización de la traza de la matriz de información, que ya había utilizado Elfving en 1952, [70]. Dos años más tarde De la Garza, [62], demuestra que para un modelo polinómico de grado  $m - 1$  y dado un diseño con más de  $m$  puntos en su soporte, siempre existe otro diseño con  $m$  puntos en su soporte y que tiene la misma matriz de información. Kiefer en 1959 corregirá esta demostración. En 1955 Ehrenfeld, [68], establece un nuevo criterio de optimización que consiste en maximizar el mínimo autovalor de la matriz de información.

En 1958 Hoel, [107], comprueba en algunos casos que los criterios de Smith y de Wald dan los mismos resultados. Con esto se muestra precursor del teorema de equivalencia dado por Kiefer y Wolfowitz en 1959, [128]. También en 1958 Guest, [96], demostró que en la regresión polinomial el diseño óptimo obtenido con el criterio de Smith tiene su soporte en los ceros de la derivada de un polinomio de Legendre.

Kiefer y Wolfowitz han contribuido en gran medida al diseño óptimo de experimentos. A ellos se deben dos grandes resultados: la idea del diseño como medida y el teorema de equivalencia. También proviene de ellos la consideración del problema de optimización parcial cuando no interesa o no es necesaria la optimización de todos los parámetros. Kiefer extiende el teorema de equivalencia a esta situación. Sugieren también en 1959, [128], el empleo de la teoría de juegos para la construcción de diseños. Sin embargo, parece más útil esta teoría para la verificación de la optimización de un diseño dado que para su construcción. Kiefer y Wilfovitz en 1960-61, [129] y [123], dan ejemplos de la construcción de diseños D-óptimos, pero no establece un método general. Fedorov y sus seguidores son los primeros en desarrollar un método general para la construcción del diseño D-óptimo. Demuestran también que dicho algoritmo converge y dan un valioso procedimiento para calcular la matriz de información y su inversa en cada paso a partir de los cálculos hechos en el paso anterior.

En 1959 G. E. P. Box y Lucas, [41], aplicaron el criterio de D-optimización en modelos no lineales. Usando un argumento geométrico obtienen diseños de  $m$  puntos para modelos de  $m$  parámetros. Demuestran que el diseño D-óptimo maximiza el volumen del simplex definido por la imagen del diseño.

En 1965 G. E. P. Box y Hunter, [40], obtienen un algoritmo para la determinación del diseño D-óptimo en el modelo no lineal. Se trata esencialmente de una aplicación de la versión de Wynn del teorema de equivalencia. Al año siguiente Karlin y Studden, [120], extienden y simplifican los resultados de optimización parcial, en particular cuando la matriz de información es singular. En 1967 Draper y Hunter, [66], discutieron el problema de seleccionar distribuciones de parámetros a priori con el objeto de obtener diseños para modelos no lineales. Un año más tarde Atkinson y Hunter, [14], extendieron los resultados de Box y Lucas al caso en que el diseño toma más de  $m$  puntos. En 1968-70 M. J. Box ([42], [43], [44] y [44]) da algunos resultados adicionales para diseños no lineales. En 1969 Atwood, [15], obtiene mejores resultados en optimización parcial, y estudia el papel de la simetría en diseño óptimo.

El cálculo del diseño óptimo no es fácil de implementar. Por este motivo Nalimov et al en 1970, [156], y Box y Draper en 1971, [46], restringen el diseño a un número fijo de observaciones  $N$ , y hablarán de DN-optimización y de GN-optimización, pero entonces no se cumplirá el teorema de equivalencia, y por tanto no se puede aplicar el algoritmo de Fedorov. Otros autores desarrollarán algoritmos para su obtención, pero sin garantizar la convergencia. En 1972 Wynn, [220], extiende el algoritmo, que él mismo había creado poco tiempo antes, a la optimización parcial de los parámetros. Está basado en la extensión del teorema de equivalencia.

Por su parte Silvey y Titterington en 1973, [187], dan una interpretación geométrica del diseño óptimo y plantean un algoritmo para obtener un diseño D-óptimo en el espacio dual. En este mismo año White, [214], extiende el teorema de equivalencia a diseños para modelos no lineales. En 1974 Kiefer, [124], da otros resultados de equivalencia para otros criterios. Cuatro años más tarde Wu y Wynn, [218], dan condiciones generales para la convergencia de los algoritmos para la obtención del diseño óptimo.

En 1980 Hill, [106], demostró que si un modelo no es lineal en alguno de los parámetros, entonces el diseño D-óptimo no depende del valor de los parámetros en que es lineal. En 1982 Currie, [60], compara diversos diseños para estimar los parámetros en la ecuación de Michaelis-Menten, frecuentemente utilizada en cinética de enzimas. Un año más tarde Abdelbasit y Plackett, [1], trabajan con modelos de regresión logística y obtienen diseños que maximizan la información sobre los parámetros en el modelo. Otros desarrollos recientes se deben a Atkinson (1982, [9]), Pázman (1980, [166]), ... El artículo de Ash y Hedayat (1978, [7]) es una amplia recopilación de la bibliografía sobre diseño óptimo hasta ese momento. Una buena introducción al tema la hacen John y Draper en 1975, [117]. Los libros de Fedorov (1972, [74]), Silvey (1980, [186]), Pázman (1986, [168]) y los recién editados de Atkinson (1992, [10]) y Pulkenseim (1993, [173]) son un buen compendio de los resultados más importantes obtenidos hasta esos momentos. En 1985 se publicó un libro recogiendo una colección de artículos de Kiefer sobre diseño óptimo de experimentos (vase [50]). Dicha colección es de un inestimable valor para los investigadores en esta materia.

El uso de la informática en los diversos campos de la estadística supone un avance considerable. En particular, en el diseño óptimo de experimentos los primeros que investigan sobre esto son G. E. P. Box y Hunter en 1965, [40], para modelos no lineales. La ayuda del ordenador fue estimulada con el fin de conseguir diseños óptimos exactos en  $N$  pruebas. El algoritmo informático más popular es DETMAX, desarrollado por Mitchell en 1974, [149], para la búsqueda de diseños D-óptimos. En 1980 Galil y Kiefer, [86], hacen algunas modificaciones. En 1982 Welch, [211], desarrolla un nuevo programa más completo. Atkinson (1992, [10]) propone un programa para diseños exactos.

Paralelamente en 1974 Snee y Marquardt, [190], desarrollan el programa XVERT para el diseño óptimo en mixturas de modelos. En 1983 Nigam y Gupta, [161], proponen una nueva versión de este algoritmo.

## 1.4. Contexto del diseño óptimo

En lo que sigue utilizaremos modelos de regresión con observaciones incorreladas. El modelo vendrá determinado por los integrantes que describimos a continuación. En primer lugar hemos de especificar el conjunto de puntos observables, donde valoran las llamadas variables controlables. Dicho conjunto recibe el nombre de *espacio del diseño* o *dominio experimental* y será denotado por  $X$ . En la práctica el espacio  $X$  va a ser un subconjunto compacto de un espacio euclídeo (con frecuencia un intervalo de la recta real). Por este motivo no constituye restricción grave suponer desde ahora en adelante que dicho conjunto es compacto.

Entenderemos por *estado*,  $\theta$ , una función que asigna a cada punto de  $X$  el promedio de las cantidades  $y(x)$  observadas en él. Estas cantidades son variables aleatorias que dependen de la inestabilidad de las condiciones, siendo su varianza conocida:

$$\sigma^2(x) = E(\{y(x) - E[y(x)]\}^2)$$

mientras que la esperanza,  $E[y(x)]$ , es desconocida. Un buen diseño tratará de reducir al mínimo la influencia de la inestabilidad de las condiciones.

El modelo de regresión puede expresarse de la forma:

$$\theta(x) = E[y(x)] = \eta(x), x \in X$$

o bien, escrito de otra manera:

$$y(x) = \eta(x) + \varepsilon(x), x \in X \text{ y siendo } E[\varepsilon(x)] = 0$$

donde la función  $\eta$  se conoce como *superficie respuesta* o *función de regresión*. Habitualmente supondremos que dicha función es parcialmente conocida, es decir, que está dentro de una familia paramétrica de funciones:

$$\eta(x) = \eta(x, \alpha)$$

donde los parámetros  $\alpha^t = (\alpha_1, \dots, \alpha_m) \in \mathfrak{R}^m$  son desconocidos y su especificación determina totalmente a  $\eta$ . Sin embargo no es la estimación de estos parámetros lo que perseguimos, sino más bien diseñar el experimento de forma que dicha estimación sea óptima.

El caso más interesante, y que trataremos especialmente es el modelo lineal con observaciones incorreladas. El problema se plantea de la forma siguiente:

$$\theta(x) = E[y(x)] = \alpha^t f(x) = \alpha_1 f_1(x) + \dots + \alpha_m f_m(x), x \in X$$

donde  $f^t(x) = (f_1(x), \dots, f_m(x))$  es una función vectorial continua y conocida.

Por tanto el *modelo de regresión lineal con observaciones incorreladas* viene determinado por una terna  $(X, \Theta, \sigma)$  donde  $X$  es un espacio métrico y compacto,  $\Theta$  es un  $\mathfrak{R}$ -espacio vectorial de dimensión finita  $m$ , compuesto por funciones reales y continuas definidas en el conjunto  $X$ . Siempre se puede conseguir que las funciones  $f_1(x), \dots, f_m(x)$  formen una base de este espacio. Basta para ello tomar una base del espacio vectorial  $\Theta$ , y redefinir el problema con nuevos parámetros. El espacio vectorial  $\Theta$  será llamado *espacio de estados*. Por su parte  $\sigma$  será una función positiva y continua. Además las observaciones  $y(x), x \in X$ , son variables aleatorias incorreladas.

Si de antemano suponemos que el número de observaciones que podemos realizar es  $N$ , llamaremos *diseño de tamaño fijo* o *diseño exacto de tamaño  $N$*  a una sucesión de  $N$  puntos de  $X$ :  $x_1, \dots, x_N$ ; donde eventualmente podrían coincidir algunos de ellos. Con el objeto de no repetir puntos denotaremos por  $N_x$  el número de observaciones realizadas en el punto  $x$ . Podemos entonces asociar a este diseño la medida discreta:

$$\xi(x) = \frac{N_x}{N}, x \in X$$

Esto sugiere una definición más general de *diseño aproximado* o *asintótico* como una medida discreta de probabilidad,  $\xi$ , en  $X$  con soporte finito.

Cabría aún una definición más general del diseño como una medida de probabilidad cualquiera, en cuyo caso suele denominarse *diseño continuo*. En la práctica no nos interesan estos dos tipos de diseño, ya que no es posible realizar infinitas observaciones o un número decimal,  $N_x = N\xi(x)$  de ellas. No obstante son convenientes para demostrar ciertas propiedades. Además eligiendo  $N$  suficientemente grande, podremos aproximar un diseño de estas características a uno exacto.

El soporte de un diseño  $\xi$  se denotará por:

$$X_\xi = \{x \in X : \xi(x) > 0\}$$

Como ya hemos apuntado más arriba, los mejores diseños serán aquéllos que en algún sentido minimicen la varianza de los estimadores de los parámetros, que son el objeto último de un buen diseño

experimental. Puesto que no podemos minimizar todas las varianzas simultáneamente, trataremos de minimizarlas en algún sentido. Cuál es el criterio más apropiado dependerá de cada caso concreto. Más adelante analizaremos los distintos criterios que se emplearán con este fin.

### Ejemplo 1.1

1. Continuando con el ejemplo del primer capítulo vamos a tratar de identificar los conceptos definidos anteriormente:

Podemos considerar como dominio experimental el conjunto  $X = \{x_1, x_2, x_3, x_4\}$ , donde los puntos observables, correspondientes a las posibles pesadas, vienen definidos de la forma:

$x_1$  = Se coloca el objeto  $A$  en un plato.

$x_2$  = Se coloca el objeto  $B$  en un plato.

$x_3$  = Se colocan los dos objetos en un plato.

$x_4$  = Se coloca un objeto en cada plato.

Se podrían considerar otras posibles pesadas haciendo la precisión de si el o los objetos se colocan en un plato o en el otro, pero esto conduce a los mismos pesos cambiados de signo. Por tanto no ser necesario considerar este caso.

El modelo lineal utilizado es el siguiente:

$$y(x) = f_1(x)P_A + f_2(x)P_B + \varepsilon(x), x \in X$$

$$\theta(x) = E_\theta[y(x)] = f_1(x)P_A + f_2(x)P_B, x \in X$$

es decir, el espacio vectorial de estados será:

$$\Theta = \{f_1P_A + f_2P_B : (P_A, P_B) \in (0, \infty)^2\}$$

donde la función  $f = (f_1, f_2)$  viene definida de la forma siguiente:

$$f(x_1) = (1, 0), f(x_2) = (0, 1), f(x_3) = (1, 1) \text{ y } f(x_4) = (1, -1)$$

La función  $f$  es continua en  $X$ , que es discreto y por tanto toda función es continua. Uno de los requerimientos del problema es que no se pueden hacer más de dos pesadas. Esto significa que estamos buscando un diseño exacto de tamaño 2. El diseño clásico tomaba la base canónica  $f(x_1) = (1, 0)$ ,  $f(x_2) = (0, 1)$ , sin repetición de ninguna de las dos pesadas, es decir, se trata del diseño:

$$\xi_1 = \left\{ \begin{array}{cc} x_1 & x_2 \\ 1/2 & 1/2 \end{array} \right\}$$

mientras que el diseño mejorado sería el siguiente:

$$\xi_2 = \left\{ \begin{array}{cc} x_3 & x_4 \\ 1/2 & 1/2 \end{array} \right\}$$

2. Establecemos ahora algunos tipos de modelos de uso frecuente en la literatura:

Modelo de observaciones directas:  $\Theta = \{\alpha_j : \alpha \in \mathfrak{R}, \text{ donde } j(x) = 1, x \in \Xi\}$ . De este modo  $E[y(x)] = \alpha + \varepsilon, x \in \Xi$ . En este modelo el interés se centra en la estimación directa de la cantidad observada, que se supone constante en todos los puntos observables de  $X$ .

Modelo de observaciones indirectas:  $\Theta = \{\sum_{i=1}^m \alpha_i f_i : (\alpha_1, \dots, \alpha_m) \in \mathfrak{R}^m\}$ , donde  $f_1, \dots, f_m$  son funciones continuas y linealmente independientes. Como hemos visto, todo modelo lineal con observaciones incorreladas puede reducirse a éste.

Modelo con emparejamiento lineal:  $X$  finito,

$$\Theta = \{\theta \in \mathcal{C}(X) : \sum_{x \in \Xi} b_i(x)\theta(x) = 0 \text{ para } i = 1, \dots, q\} \quad (1.1)$$

siendo  $b_1, \dots, b_q$  funciones linealmente independientes. Con este modelo se busca discriminar cuáles de las funciones que verifican las condiciones impuestas en (1.1) son las que mejor se ajustan a los datos observados. Un ejemplo sería el siguiente:

$$\Theta = \{\theta \in \mathcal{C}([-1, 1]) : \theta(-1) = \theta(1) \text{ y } \theta(0) = 0\}$$

En este caso se está imponiendo la condición de que las funciones a ajustar tomen valores idénticos en los extremos y que pasen por el origen. Estas restricciones pueden ser datos que se conocen a priori y que se muestran convenientes para obtener un ajuste adecuado.

Modelo de observaciones indirectas bajo restricciones lineales:

$$\Theta = \left\{ \sum_{i=1}^m \alpha_i f_i : \sum_{j=1}^m B_{kj} \alpha_j = 0 \text{ para } k = 1, \dots, q \right\}$$

siendo  $B_{kj}$  números dados tales que el rango de la matriz:

$$\begin{pmatrix} B_{11} & \dots & B_{1m} \\ \dots & \dots & \dots \\ B_{q1} & \dots & B_{qm} \end{pmatrix}$$

sea máximo. Este modelo se podría reducir a un modelo lineal de observaciones indirectas reduciendo el número de parámetros a  $m - q$ .

Modelo de regresión spline: Sea  $X = [a, b]$  y sea una partición del intervalo  $x_0 = a < x_1 < \dots < x_k = b$ . Supongamos que  $f_1^{(i)}, \dots, f_m^{(i)}$  son funciones linealmente independientes en el intervalo  $[x_{i-1}, x_i], i = 1, \dots, k$ . El espacio de estados será:

$$\Theta = \left\{ \theta \in \mathcal{C}(X) : \theta_{[x_{i-1}, x_i]} \in \langle f_1^{(i)}, \dots, f_m^{(i)} \rangle, \right. \\ \left. \lim_{x \rightarrow x_i^+} \frac{\partial^j \theta(x)}{\partial x^j} = \lim_{x \rightarrow x_i^-} \frac{\partial^j \theta(x)}{\partial x^j}, i = 1, \dots, k-1, j = 0, \dots, r \right\}$$

Es un modelo de regresión a trozos que conecta con suavidad en los extremos de los intervalos de la partición. Para  $j = 0$  tenemos la condición de continuidad, mientras que para  $j = 1$  tenemos la de diferenciabilidad. Así, en el intervalo  $[x_{i-1}, x_i]$  el modelo será de la forma:

$$\alpha_1^{(i)} f_1^{(i)} + \dots + \alpha_m^{(i)} f_m^{(i)}$$

A modo de ejemplo supongamos que  $X = [0, 2] = [0, 1] \cup [1, 2]$ , y que el modelo es  $E[y(x)] = \alpha_1 + \alpha_2 x$  en  $X = [0, 1]$  y  $E[y(x)] = \beta_1 x + \beta_2 x^3$  en  $X = [1, 2]$  para  $j=1$ . Se tendrán las siguientes condiciones de unión:

$$\alpha_2 = \beta_1 + 3\beta_2$$

$$\alpha_1 + \alpha_2 = \beta_1 + \beta_2$$

y habrá que estimar solamente dos parámetros.

Modelo del Análisis de la Varianza con dos factores:

$$X = \{(i, j) : i \in \{1, \dots, v\}, j \in \{1, \dots, b\}\}$$

$$\Theta = \{\theta(i, j) = \mu + \alpha_i + \beta_j : \sum \alpha_i = \sum \beta_j = 0\}$$

Cada punto observable determina en qué categoría se ha de introducir el individuo observado. El valor observado dependerá, en principio, de los dos factores, que pueden tomar cada uno  $v$  y  $b$  valores respectivamente. No se considera interacción entre los factores.

Este caso podrá reducirse al modelo de observaciones indirectas con restricciones lineales.



## Capítulo 2

# Estimadores de funcionales lineales

En muchas ocasiones va a resultar más interesante estimar determinadas relaciones entre los parámetros que estimar cada uno de ellos. Esto sugiere la siguiente definición de *funcional lineal*,  $g$ , como una función lineal del espacio de estados en la recta real:

$$g : \Theta \longrightarrow \Re$$

Supondremos que la aplicación  $g$  tiene como matriz asociada en la base  $\{f_1, \dots, f_m\}$  de  $\Theta$  el vector  $c^t = (c_1, \dots, c_m)$ . Es decir, dado un estado:

$$\theta(x) = \sum_{i=1}^m \alpha_i f_i(x)$$

entonces  $g(\theta) = \alpha^t c$ .

Buscaremos entonces una buena estimación de  $g(\theta)$ . Su valor será calculado a partir de los datos experimentales, utilizando para ello una función lineal de las variables respuesta:

$$\sum_{i=1}^N a_i y(x_i)$$

donde  $a_1, \dots, a_N$  son ciertos coeficientes que habrá que determinar bajo ciertas exigencias. A esta función se le llamará *estimador lineal* de  $g$ . Diremos además que es un *estimador centrado* siempre que:

$$E\left[\sum_{i=1}^N a_i y(x_i)\right] = g(\theta)$$

Diremos que  $g$  es *estimable* si existe al menos un estimador lineal centrado de  $g$ . El mejor estimador lineal centrado será aquél que tenga mínima varianza (BLUE, the best linear unbiased estimator), es decir, el estimador:

$$\sum_{i=1}^N a_i^* y(x_i)$$

tal que:

$$\text{var}\left[\sum_{i=1}^N a_i^* y(x_i)\right] = \min\left\{\text{var}\left[\sum_{i=1}^N a_i y(x_i)\right] : (a_1, \dots, a_N) \in \mathfrak{R}^N, E\left[\sum_{i=1}^N a_i y(x_i)\right] = g(\theta), \theta \in \Theta\right\}$$

es decir, dicho estimador vendrá dado por los coeficientes  $a_i^*$  tales que:

$$\sum_{i=1}^N a_i^{*2} \sigma_i^2 = \min\left\{\sum_{i=1}^N c_i^2 \sigma_i^2 : (a_1, \dots, a_N) \in \mathfrak{R}^N, \sum_{i=1}^N a_i \theta(x_i) = g(\theta), \theta \in \Theta\right\}.$$

Nos interesa obtener una expresión explícita de estos coeficientes. A ello, como un primer paso, va encaminada la definición que daremos a continuación. Dado un diseño de tamaño fijo  $N$ , utilizaremos la siguiente notación:

$$\begin{aligned} Y &= (y(x_1), \dots, y(x_N))^t \\ C &= (a_1, \dots, a_N)^t \\ X &= \begin{pmatrix} f_1(x_1) & \dots & f_m(x_1) \\ \dots & \dots & \dots \\ f_1(x_N) & \dots & f_m(x_N) \end{pmatrix} \end{aligned}$$

**Definición 2.1** La matriz de información de un diseño exacto  $x_1, \dots, x_N$ , se define como la matriz:

$$M = X^t \Sigma^{-1} X$$

donde  $\Sigma = \text{diag}(\sigma^2(x_1), \dots, \sigma^2(x_N))$ .

Por construcción esta matriz es simétrica y semidefinida positiva:

$$u^t M u = u^t X^t \Sigma^{-1} X u = (\Sigma^{-1/2} X u)^t (\Sigma^{-1/2} X u) = \|\Sigma^{-1/2} X u\|^2 \geq 0, u \in \mathfrak{R}^m$$

Teniendo en cuenta que esta matriz es la inversa de la matriz de covarianzas, se podrá haber concluido que es semidefinida positiva.

Hasta ahora hemos definido la matriz de información y hemos calculado los estimadores utilizando diseños exactos. Esto lo generalizaremos a continuación para diseños aproximados o asintóticos.

**Definición 2.2** Se define la matriz de información asociada a un diseño aproximado  $\xi$  como la matriz de orden  $m$ :

$$M(\xi) = \sum_{x \in X} f(x) f^t(x) \sigma^{-2}(x) \xi(x)$$

**Observación 2.1** Puede darse una definición análoga en cuanto a la matriz de información de un diseño continuo. A partir de ahora utilizaremos solamente diseños aproximados o asintóticos y los denominaremos simplemente diseños. Si  $\xi$  es un diseño exacto  $x_1, \dots, x_N$  entonces según esta nueva definición de matriz de información asociada a un diseño en general tendremos:

$$M(\xi) = \frac{1}{N} \sum_{x \in X} f_i(x) f_j(x) \sigma^{-2}(x) N_x = \frac{1}{N} M$$

En muchas ocasiones suponer que la varianza de las observaciones es uno,  $\sigma^2(x) = 1$ , no supone una pérdida de generalidad, sin más que sustituir  $\sigma^{-1}(x)\theta(x)$  y  $\sigma^{-1}(x)f(x)$  por  $\theta(x)$  y  $f(x)$  respectivamente. De este modo el error sigue teniendo esperanza nula y la varianza del nuevo modelo se hace igual a uno. En este caso la matriz de información quedará:

$$M(\xi) = \sum_{x \in X} f(x)f^t(x)\xi(x)$$

En lo que sigue se supondrá siempre que  $\sigma^2(x) = 1$ , salvo que se especifique lo contrario.

### Ejemplo 2.1

Continuando con el modelo del ejemplo estudiado en el primer capítulo y con los dos diseños propuestos, calculamos ahora las matrices de información asociadas a ellos:

$$M(\xi_1) = \frac{\sigma^{-2}}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1, 0) + \frac{\sigma^{-2}}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0, 1) = \begin{pmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}$$

$$M(\xi_2) = \frac{\sigma^{-2}}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1, 1) + \frac{\sigma^{-2}}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} (1, -1) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{pmatrix}$$

Supongamos ahora que estamos interesados en estimar la suma de los pesos de los dos objetos, es decir, el funcional lineal  $g(\theta) = P_A + P_B$ , de modo que  $c = (1, 1)$ . Puesto que tanto los estimadores de  $P_A$  y  $P_B$  obtenidos a través del diseño clásico como los obtenidos a través del diseño mejorado eran centrados entonces  $\hat{P}_A + \hat{P}_B$  será un estimador centrado de  $g$ . Además dicho estimador es combinación lineal de  $y_0, \dots, y_4$  en ambos casos. Se deduce por tanto que  $g$  es estimable. Buscaremos el BLUE. Para ello hemos de minimizar la función:

$$\begin{cases} \psi(a_1, \dots, a_4) = \sum_{i=0}^4 c_i^2 \sigma^2 \\ \text{con la restricción } \sum_{i=0}^4 a_i \theta(x_i) = g(\theta) \end{cases}$$

Puesto que  $\theta(x_1) = P_A$ ,  $\theta(x_2) = P_B$ ,  $\theta(x_3) = P_A + P_B$  y  $\theta(x_4) = P_A - P_B$ , la restricción se reduce a  $a_1 P_A + a_2 P_B + a_3 (P_A + P_B) + a_4 (P_A - P_B) = P_A + P_B$ . Puesto que ha de cumplirse para todo posible valor de los parámetros  $P_A$  y  $P_B$  igualando los coeficientes de estos tendremos las dos restricciones correspondientes que dan lugar al problema de programación cuadrática con restricciones lineales siguiente:

$$\begin{cases} \text{mín } \sum_{i=0}^4 c_i^2 \\ \text{con } \begin{cases} 1 - a_1 - a_3 - a_4 = 0 \\ 1 - a_2 - a_3 + a_4 = 0 \end{cases} \end{cases}$$

Resolviendo este problema con los multiplicadores de Lagrange, obtenemos la solución siguiente:

$$a_1 = 1/3, a_2 = 1/3, a_3 = 2/3, a_4 = 0$$

de modo que el BLUE de  $g$  es:

$$\hat{g} = y_1/3 + y_2/3 + 2y_3/3$$

y su varianza:

$$\text{var}(\hat{g}) = 2\sigma^2/3$$

**Notación 2.1** Sea  $A$  una matriz cualquiera de orden  $m$ . Se definen los conjuntos siguientes:

$$\mathcal{M}(A) = \{Au : u \in \mathfrak{R}^m\}$$

$$\mathcal{N}(A) = \{u \in \mathfrak{R}^m : Au = 0\}$$

que son los subespacios imagen y núcleo de la aplicación lineal asociada a la matriz  $A$ .

El conjunto de todos los diseños en el modelo se denotará por  $\Xi$ , mientras que el conjunto de todas las matrices de información será:

$$\mathcal{M} = \{M(\xi) : \xi \in \Xi\}$$

El conjunto  $\mathcal{M}$  tiene en general una estructura más sencilla que el conjunto  $\Xi$ . De hecho, como veremos en la siguiente proposición,  $\mathcal{M}$  es un subconjunto convexo de un espacio euclídeo, el de las matrices cuadradas de orden  $m$  y simétricas. Además la varianza de un estimador será función de la  $g$ -inversa de la matriz de información.

**Proposición 2.1** El conjunto  $\mathcal{M}$  es convexo.

**Demostración:** Sean  $\xi_1, \xi_2 \in \Xi$  y  $0 < \lambda < 1$ , entonces:

$$(1 - \lambda)M(\xi_1) + \lambda M(\xi_2) = M[(1 - \lambda)\xi_1 + \lambda\xi_2] \in \mathcal{M}$$

**Proposición 2.2** 1. La matriz de información es simétrica y semidefinida positiva.

2. Si  $\xi$  tiene menos de  $m$  puntos en su soporte, entonces  $\det M(\xi) = 0$ .

**Demostración:**

1. La definición de la matriz de información muestra directamente que se trata de una matriz simétrica. Además, sea  $u$  vector de  $\mathfrak{R}^m$ , entonces:

$$u^t M(\xi) u = \sum_{x \in X} u^t f(x) f^t(x) u \sigma^{-2}(x) \xi(x) = \sum_{x \in X} \|u^t f(x)\|^2 \sigma^{-2}(x) \xi(x) \geq 0$$

2. Supongamos que  $\xi$  tiene en su soporte  $k < m$  puntos:  $x_1, \dots, x_k$ . En el desarrollo del determinante aparecerán siempre al menos dos columnas iguales y por tanto el determinante ha de ser cero.

**Observación 2.2** Deduciremos ahora una expresión explícita del determinante de la matriz de información. Para simplificar la notación supondremos que  $\sigma(x) = 1$ . En primer lugar trataremos el caso en el que el soporte se reduce a  $m$  puntos,  $X_\xi = \{x_1, \dots, x_m\}$ . Llamaremos:

$$m_{ij} = \sum_{x \in X} f_i(x) f_j(x) \xi(x)$$

$$a_{ik} = f_i(x_k) \xi(x_k), b_{jk} = f_j(x_k)$$

Denotando  $A = (a_{ik})$  y  $B = (b_{jk})$  tendremos la siguiente expresión:

$$\det M(\xi) = \det A \det B = \prod_{k=1}^m \xi(x_k) \det[f_i(x_j)]^2 \quad (2.1)$$

En el caso más general en que  $X_\xi = \{x_1, \dots, x_r\}$  con  $r \geq m$  tendremos:

$$\begin{aligned} \det M(\xi) &= \det[\{\sum_{k=1}^r f_i(x_k) f_j(x_k) \xi(x_k)\}] = \sum_{\tau \in S_r} \det[\{f_i(x_{\tau(j)}) f_j(x_{\tau(j)}) \xi(x_{\tau(j)})\}] = \\ &= \sum_{\tau \in S_r} \prod_{j=1}^m \xi(x_{\tau(j)}) \det[\{f_i(x_{\tau(j)})\}]^2 = \sum_{k_1 < \dots < k_m} \xi(x_{k_1}) \cdots \xi(x_{k_m}) \det[\{f_i(x_{k_j})\}]^2 \end{aligned}$$

donde hemos llamado  $S_r$  al grupo simétrico de las permutaciones de orden  $r$ . La tercera igualdad se ha obtenido a partir de (2.1).

### Ejemplo 2.2

Sea el modelo con dos parámetros  $\theta(x) = \alpha f(x) + \beta g(x)$ ,  $\sigma^2(x) = 1$ ,  $x \in X$ . El determinante de la matriz de información asociada a un diseño cualquiera será:

$$\begin{aligned} \det M(\xi) &= \sum_{i < j} \xi(x_i) \xi(x_j) \det \begin{pmatrix} f(x_i) & f(x_j) \\ g(x_i) & g(x_j) \end{pmatrix}^2 = \\ &= \sum_{i < j} \xi(x_i) \xi(x_j) [f(x_i)g(x_j) - f(x_j)g(x_i)]^2 \end{aligned}$$

**Observación 2.3** La siguiente proposición nos dará una expresión explícita de la varianza del BLUE del funcional  $g$ . Para ello se hará uso de la inversa generalizada de una matriz.

Dada una matriz cualquiera,  $A$ , diremos que  $A^-$  es una inversa generalizada o  $g$ -inversa cuando  $AA^-A = A$ . Otra definición equivalente a ésta es la siguiente:  $A^-$  es una inversa generalizada de  $A$  si  $A^-u$  satisface la ecuación  $Ax = u$  para cada  $u$  de  $\mathcal{M}(A)$ . La inversa generalizada existe siempre, pero en general no es única. Si  $A$  es cuadrada y regular entonces la inversa generalizada coincide con la matriz inversa.

Cuando la matriz  $A$  es simétrica, existe una  $g$ -inversa muy particular, debida a Penrose, y que denotaremos por  $A^+$ . Esta matriz es la única que verifica lo siguiente:

$$A^+u = \begin{cases} 0 & \text{si } u \in \mathcal{N}(A) \\ w_u & \text{si } u \in \mathcal{M}(A) \end{cases}$$

donde  $w_u$  es el único vector de  $\mathcal{M}(A)$  tal que  $Aw_u = u$ .

**Ejercicios.-** Demostrar que:

1. Las dos definiciones de inversa generalizada son equivalentes.
2. Siempre existe la inversa generalizada.
3. Dar un ejemplo en el que la inversa generalizada no sea única.
4. Si  $A$  es simétrica entonces existe un único vector  $w_u$  de  $\mathcal{M}(A)$  tal que  $Aw_u = u$ .

5.  $A^+$  es una g-inversa de  $A$  y además  $A^+AA^+ = A^+$ .

**Proposición 2.3** 1.  $g$  es estimable para el diseño  $\xi$  si, y sólo si  $c \in \mathcal{M}[M(\xi)]$ . Entonces existe un único vector  $z_c \in \mathcal{M}[M(\xi)]$  de modo que  $c = M(\xi)z_c$ . La varianza del BLUE es:

$$\begin{aligned} \text{var}(\hat{g}) &= N^{-1}z_c^t M(\xi)z_c = N^{-1}c^t M^{-}(\xi)c = N^{-1} \sup\left\{\frac{(c^t\alpha)^2}{\alpha^t M(\xi)\alpha} : \alpha \in \mathfrak{R}^m, M(\xi)\alpha \neq 0\right\} = \\ &= N^{-1} \sup\{2c^t\alpha - \alpha^t M(\xi)\alpha : \alpha \in \mathfrak{R}^m\} \end{aligned}$$

2. Si  $g_1$  y  $g_2$  son funcionales estimables, entonces:

$$\text{cov}(\hat{g}_1, \hat{g}_2) = N^{-1}c_1^t M^{-}(\xi)c_2$$

**Demostración.-** Basta tener en cuenta que:

$$\begin{aligned} \Sigma_{c^t\hat{\beta}} &= c^t \Sigma_{\hat{\beta}} c \\ \Sigma_{(c_1, c_2)^t\hat{\beta}} &= (c_1, c_2)^t \Sigma_{\hat{\beta}} (c_1, c_2) = \begin{pmatrix} c_1^t \Sigma_{\hat{\beta}} c_1 & c_1^t \Sigma_{\hat{\beta}} c_2 \\ c_2^t \Sigma_{\hat{\beta}} c_1 & c_2^t \Sigma_{\hat{\beta}} c_2 \end{pmatrix} \end{aligned}$$

**Observación 2.4** La definición general de funcional estimable hacía referencia a un diseño exacto. Así ha de entenderse la proposición anterior. Estos resultados sugieren la siguiente definición:

**Definición 2.3** Dado un diseño  $\xi$  y un funcional  $g$ , se define la varianza generalizada de  $g$  respecto del diseño  $\xi$  como:

$$\text{var}_{\xi}g = \sup\left\{\frac{(c^t\alpha)^2}{\alpha^t M(\xi)\alpha} : \alpha \in \mathfrak{R}^m, M(\xi)\alpha \neq 0\right\}$$

que puede también escribirse de la forma:

$$\text{var}_{\xi}g = \begin{cases} c^t M^{-}(\xi)c & \text{si } c \in \mathcal{M}[M(\xi)] \\ \infty & \text{si } c \notin \mathcal{M}[M(\xi)] \end{cases}$$

y se define la covarianza de dos funcionales  $g_1$  y  $g_2$  respecto de  $\xi$  como:

$$\text{cov}_{\xi}(g_1, g_2) = c_1^t M^{-}(\xi)c_2$$

Estamos interesados en buscar  $\xi$  de manera que se haga mínima esta varianza. A este fin van encaminadas las siguientes proposiciones, cuya demostración se omite.

**Proposición 2.4** Si  $M(\xi) \geq M(\eta)$ , es decir  $M(\xi) - M(\eta)$  es semidefinida positiva, entonces  $\text{var}_{\xi}g \leq \text{var}_{\eta}g$  cualquiera que sea el funcional lineal  $g$  definido en el espacio de estados. Recíprocamente, si  $\text{var}_{\xi}g \leq \text{var}_{\eta}g$  entonces se cumple que:

$$\mathcal{M}[M(\xi)] \supset \mathcal{M}[M(\eta)] \text{ y } u^t M(\xi)u \geq u^t M(\eta)u, u \in \mathcal{M}[M(\eta)]$$

**Proposición 2.5**  $M(\xi) = M(\eta)$  si, y sólo si  $\text{var}_{\xi}g = \text{var}_{\eta}g$ , para cada funcional lineal,  $g$ , definido en el espacio de estados.

**Proposición 2.6** Si  $M(\xi)$  y  $M(\eta)$  son regulares, entonces:

$$M(\xi) \geq M(\eta) \implies \text{var}_\xi g \leq \text{var}_\eta g, \text{ para cada funcional } g$$

Además:

$$M(\xi) > M(\eta) \iff \text{var}_\xi g < \text{var}_\eta g, \text{ para cada funcional } g \neq 0$$

**Proposición 2.7** Si  $\lambda_1(\xi) \leq \dots \leq \lambda_m(\xi)$  son los autovalores de  $M(\xi)$ , entonces:

$$\sum_{i=1}^k \lambda_i(\xi) \geq \sum_{i=1}^k \lambda_i(\eta), k = 1, \dots, m \iff \text{var}_\xi g \leq \text{var}_\eta g, \text{ para cada funcional } g$$

$$\sum_{i=1}^k \lambda_i(\xi) > \sum_{i=1}^k \lambda_i(\eta), k = 1, \dots, m \iff \text{var}_\xi g < \text{var}_\eta g, \text{ para cada funcional } g \neq 0$$

**Teorema 2.1 (Caratheodory)** Sea  $T$  un subconjunto de un espacio euclídeo de dimensión  $k$ . Todo punto del cierre convexo de  $T$ :

$$Co(T) = \left\{ z = \sum_{i=1}^{n(z)} \beta_i t_i : \beta_i \in [0, 1], \sum_{i=1}^{n(z)} \beta_i = 1, t_i \in T \right\}$$

puede ser expresado como una combinación convexa de a lo sumo  $k + 1$  puntos del conjunto  $T$ . Es decir, dado un elemento  $q$  de  $Co(T)$ , existirán

$$t_1, \dots, t_{k+1} \in T \text{ y } \gamma_1, \dots, \gamma_{k+1} \in [0, 1] \text{ con } \sum_{i=1}^{k+1} \gamma_i = 1$$

tal que:

$$q = \sum_{i=1}^{k+1} \gamma_i t_i$$

Este resultado puede verse fácilmente en el siguiente ejemplo donde  $T$  es un conjunto de 4 puntos no alineados en el plano. En este caso cualquier punto de  $Co(T)$  se encuentra en un triángulo formado por tres de ellos, como se muestra en la Figura 2

**Proposición 2.8** Si  $T$  es un conjunto acotado de  $\mathbb{R}^k$  entonces  $Co(T)$  es compacto en  $\mathbb{R}^k$ .

**Corolario 2.1**  $\mathcal{M}$  es compacto.

**Demostración:** Definimos:

$$\mathcal{S} = \{f(x)f^t(x) : x \in X\}$$

que es imagen de  $X$  por la aplicación continua:

$$X \longrightarrow \mathbb{R}^{m \times m} \mid x \longrightarrow f(x)f^t(x)$$

y por tanto  $\mathcal{S}$  es también compacto. Por la Proposición anterior  $\mathcal{M} = Co(\mathcal{S})$  también lo será.

Damos a continuación un resultado, muy interesante en cuanto a la búsqueda del diseño óptimo más sencillo, basado en el teorema de Caratheodory, y que es debido a Karlin S. y Studden W. I. ([120], pg 787).

Figura 2.1: Teorema de Caratheodory

**Proposición 2.9** *Dado un diseño cualquiera  $\xi$  existe otro diseño  $\eta$  tal que  $M(\xi) = M(\eta)$  y cuyo soporte tiene a lo sumo  $\frac{m(m+1)}{2} + 1$  puntos.*

**Demostración:** Definimos el vector:

$$a(x) = (f_i(x)f_j(x) : 1 \leq i \leq j \leq m)$$

A todo elemento de  $Co\{a(x) : x \in X\}$  se le puede asociar una matriz de información unívocamente. Por el teorema de Caratheodory para el diseño  $\xi$  existirá entonces:

$$\gamma_k \in [0, 1] \text{ y } x_k \in X, k = 1, \dots, \frac{m(m+1)}{2} + 1$$

de modo que definiendo:

$$\xi = \left\{ \begin{array}{ccc} x_1 & x_2 & \cdots \\ \gamma_1 & \gamma_2 & \cdots \end{array} \right\}$$

se obtiene la matriz de informacin:

$$M(\xi) = \sum_{k=1}^{\frac{m(m+1)}{2} + 1} \gamma_k f(x_k) f^t(x_k)$$

### Ejemplo 2.3

Siguiendo con el ejemplo 2.1 veamos que la varianza del BLUE de  $g$  calculada más arriba puede obtenerse a partir de la fórmula dada en la proposición 2.3 para el diseño uniforme concentrado en  $x_1, \dots, x_4$ :

$$\text{var}(\hat{g}) = N^{-1} z_g^t M(\xi) z_g = N^{-1} c^t M^{-1}(\xi) c = \frac{1}{5} (1, 1) M^{-1}(\xi) \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Hemos de calcular una matriz g-inversa de la matriz de información asociada al diseño:

$$\xi_0 = \left\{ \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{array} \right\}$$

$$M(\xi_0) = \frac{1}{4\sigma^2} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1,0) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0,1) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1,1) + \begin{pmatrix} -1 \\ -1 \end{pmatrix} (-1,-1) \right\} = \frac{1}{4\sigma^2} \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

$$M^{-1}(\xi_0) = \frac{4\sigma^2}{3} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

de modo que:

$$\text{var}(\hat{g}) = \frac{\sigma^2}{3} (1,1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{2\sigma^2}{3}$$

Además en la proposición II.4 de [168] se da una fórmula para el cálculo del BLUE de  $g$ , que hemos calculado más arriba:

$$\hat{g} = \sum_{i=1}^3 N c^t M^{-1}(\xi_0) f(x_i) \sigma^{-2}(x_i) y_i = \frac{\sigma^2}{3} (1,1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} f(x_i) \sigma^{-2} y_i = y_1/3 + y_2/3 + 2y_3/3$$

La varianza generalizada sería:

$$\text{var}_{\xi_0}(g) = c^t M^{-1}(\xi_0) c = \frac{5\sigma^2}{3} (1,1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{10\sigma^2}{3}$$

Veamos ahora que  $M(\xi_1) < M(\xi_2)$ . En efecto, la matriz:

$$M(\xi_2) - M(\xi_1) = \begin{pmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}$$

es definida positiva. Podemos deducir de aquí que  $\text{var}_{\xi_2}(g) < \text{var}_{\xi_1}(g)$ , que se comprueba haciendo los cálculos:

$$\text{var}_{\xi_2}(g) = (1,1) \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2\sigma^2 < \text{var}_{\xi_1}(g) = (1,1) \begin{pmatrix} 2\sigma^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 4\sigma^2$$

En este ejemplo buscamos el estimador de mnima varianza suponiendo que podemos tomar observaciones en los cinco puntos. No hemos considerado todava el problema de encontrar el diseo que produce mnima varianza.



## Capítulo 3

# El modelo no lineal

Hasta ahora hemos supuesto que el espacio de estados  $\Theta$  tenía estructura de espacio vectorial. Es decir, que dada una base  $\{f_1(x), \dots, f_m(x)\}$  de este espacio, todo estado podía ponerse de la forma  $\theta = \alpha^t f$ , donde los parámetros  $\alpha$  estaban unívocamente determinados para cada estado.

Sin embargo en la práctica nos encontraremos con frecuentes problemas de regresión no lineal. Si podemos seguir considerando funcionales lineales  $g$ , en cierto sentido:

$$g(\beta_1\theta_1 + \beta_2\theta_2) = \beta_1g(\theta_1) + \beta_2g(\theta_2)$$

siempre que:

$$\theta_1, \theta_2 \in \Theta \text{ y } \beta_1\theta_1 + \beta_2\theta_2 \in \Theta$$

bastará extender  $g$  linealmente al espacio vectorial  $\mathcal{L}(\Theta)$  generado por  $\Theta$ . Denotaremos dicha extensión por  $\tilde{g}$ . De este modo estamos transformando el problema en un modelo de regresión lineal  $(X, \mathcal{L}(\Theta), \sigma)$ . Se cumple entonces que un estimador es centrado para  $g$  si, y sólo si lo es para  $\tilde{g}$ .

Pero si nos interesa estimar funcionales no lineales, la solución no es tan sencilla. Entenderemos entonces por *modelo de regresión no lineal* una terna  $(X, \Theta, \sigma)$  donde:

$$\Theta = \{\eta(\cdot, \alpha) : \alpha \in U\}$$

siendo  $U$  un abierto de  $\mathfrak{R}^m$ , y  $\eta$  una función real definida en  $X \times U$ , continua en  $X$  y que es no lineal, pero diferenciable, en los parámetros  $\alpha_1, \dots, \alpha_m$ . No podemos ahora definir la matriz de información como hacíamos antes. Pero si utilizamos la fórmula de Taylor para linealizar estas funciones, transformaremos el modelo no lineal en uno lineal. Para ello necesitaremos hacer una estimación inicial de los parámetros,  $\alpha_1^{(0)}, \dots, \alpha_m^{(0)}$ . El desarrollo de Taylor de segundo grado en un entorno de  $\alpha^{(0)}$  quedará de la forma:

$$\eta(x, \alpha) = \eta(x, \alpha^{(0)}) + \sum_{i=1}^m \left( \frac{\partial \eta(x, \alpha)}{\partial \alpha_i} \right)_{\alpha^{(0)}} (\alpha_i - \alpha_i^{(0)}) + \frac{1}{2} \sum_{i,j=1}^m \left( \frac{\partial^2 \eta(x, \alpha)}{\partial \alpha_i \partial \alpha_j} \right)_{\alpha^{(s)}} (\alpha_i - \alpha_i^{(0)}) (\alpha_j - \alpha_j^{(0)})$$

con  $\alpha^{(s)}$  situado en el segmento que une los puntos  $\alpha^{(0)}$  y  $\alpha$ . Si hacemos  $\beta_i = \alpha_i - \alpha_i^{(0)}$ ,  $i = 1, 2, \dots, m$ ; entonces  $\theta(x) = \eta(x, \alpha) - \eta(x, \alpha^{(0)})$  y:

$$\tilde{f}_i(x) = \left( \frac{\partial \eta(x, \alpha)}{\partial \alpha_i} \right)_{\alpha^{(0)}} ; i = 1, \dots, m$$

entonces el problema puede aproximarse por los siguientes estados:

$$\theta(x) = \sum_{i=1}^m \tilde{f}_i(x) \beta_i$$

Pero ahora la estimación de los parámetros será sesgada. El sesgo viene introducido por el resto de la fórmula de Taylor:

$$R_x(\alpha^{(0)}, \alpha^{(s)}, \alpha) = \frac{1}{2} \sum_{i,j=1}^m m \left( \frac{\partial^2 \eta(x, \alpha)}{\partial \alpha_i \partial \alpha_j} \right)_{\alpha^{(s)}} (\alpha_i - \alpha_i^{(0)}) (\alpha_j - \alpha_j^{(0)})$$

En [168] se demuestra que el BLUE de  $\beta$  es:

$$\hat{\beta} = M^{-1}(\xi) \sum_{x \in X} \tilde{f}(x) [y(x) - \eta(x, \alpha^{(0)})] \xi(x)$$

siendo este estimador sesgado en el modelo original. Su sesgo viene expresado por:

$$E(\hat{\beta}) - \beta = M^{-1}(\xi) \sum_{x \in X} \tilde{f}(x) R_x(\alpha^{(0)}, \alpha^{(s)}, \alpha^*) \xi(x)$$

donde  $\alpha_i^*$  son ciertos números comprendidos entre  $\alpha_i^{(0)}$  y  $\alpha_i^{(s)}$ ,  $i = 1, \dots, m$ . En algunas ocasiones no es viable la linealización por ser muy tosca la aproximación obtenida por Taylor. Entonces deberán utilizarse otros métodos.

### Ejemplo 3.1

1. Sea el espacio de estados  $\Theta = \{\theta = \alpha f(x) + \beta g(x) : \alpha^2 + \beta^2 \leq 1\}$  con  $f$  y  $g$  linealmente independientes, y sea la función:

$$g : \Theta \longrightarrow \Re \mid \theta \longrightarrow \alpha + \beta$$

Se trata de una función lineal en cierto sentido, es decir, cuando  $\theta_1, \theta_2, \lambda_1 \theta_1 + \lambda_2 \theta_2 \in \Theta$ ; o lo que es igual:

$$\begin{aligned} \alpha_i^2 + \beta_i^2 &\leq 1, i = 1, 2 \\ (\lambda_1 \alpha_1 + \lambda_2 \alpha_2)^2 + (\lambda_1 \beta_1 + \lambda_2 \beta_2)^2 &\leq 1 \end{aligned}$$

Considerando el espacio vectorial generado por  $\Theta$ , entonces  $g$  puede extenderse a un funcional lineal de dicho espacio:

$$\tilde{g} : \mathcal{L}(\Theta) = \{\theta = \alpha f + \beta g\} \longrightarrow \Re \mid \theta \longrightarrow \alpha + \beta$$

Además, como ya hemos dicho, si:

$$E\left[\sum_{i=1}^N a_i y(x_i)\right] = \sum_{i=1}^N a_i E[y(x_i)] = \sum_{i=1}^N a_i [\alpha f(x_i) + \beta g(x_i)] = \alpha + \beta, \text{ cuando } \alpha^2 + \beta^2 \leq 1$$

entonces se cumple también para todo  $\alpha$  y  $\beta$ .

2. Supongamos que la función de superficie es no lineal, del tipo:

$$\eta(x, \alpha, \beta) = \alpha_1 e^{\beta_1 x} + \dots + \alpha_m e^{\beta_m x}, x \in [0, 1]$$

y utilizando el desarrollo de Taylor de primer orden tendremos:

$$\begin{aligned} \eta(x, \alpha, \beta) = & \eta(x, \alpha^{(0)}, \beta^{(0)}) + \sum_{i=1}^m \left( \frac{\partial \eta}{\partial \alpha_i} \right)_{(\alpha^{(0)}, \beta^{(0)})} (\alpha_i - \alpha_i^{(0)}) + \\ & \sum_{i=1}^m \left( \frac{\partial \eta}{\partial \beta_i} \right)_{(\alpha^{(0)}, \beta^{(0)})} (\beta_i - \beta_i^{(0)}) + R_x(\alpha_i, \alpha_i^{(0)}, \alpha_i^{(s)}) \end{aligned}$$

y llamando  $\gamma_i = \alpha_i - \alpha_i^{(0)}$ ,  $\delta_i = \beta_i - \beta_i^{(0)}$ ,  $\tilde{f}_i(x) = \left( \frac{\partial \eta}{\partial \alpha_i} \right)_{(\alpha^{(0)}, \beta^{(0)})} = e^{\beta_i^{(0)} x}$  y  $\tilde{g}_i(x) = \left( \frac{\partial \eta}{\partial \beta_i} \right)_{(\alpha^{(0)}, \beta^{(0)})} = \alpha_i^{(0)} x e^{\beta_i^{(0)} x}$ , podremos expresar el modelo de la forma:

$$\begin{aligned} \theta(x) = \eta(x, \alpha, \beta) - \eta(x, \alpha^{(0)}, \beta^{(0)}) & \approx \alpha^t \tilde{f}(x) + \beta^t \tilde{g}(x) = \\ \alpha_1 e^{\beta_1^{(0)} x} + \dots + \alpha_m e^{\beta_m^{(0)} x} + \beta_1 \alpha_1^{(0)} x e^{\beta_1^{(0)} x} + \dots + \beta_m \alpha_m^{(0)} x e^{\beta_m^{(0)} x} \end{aligned}$$



## Capítulo 4

# Criterios de optimización

### 4.1. Introducción

¿Qué entendemos por diseño óptimo de un experimento? ¿Cómo entender la expresión “el mejor de los diseños posibles”? Estas preguntas no tienen una respuesta fácil. Si bien es verdad que nos interesará el diseño que haga mínima la varianza, también es cierto que un diseño puede hacer mínima la varianza para un funcional lineal, y excesivamente grande para otro.

Necesitamos por tanto elegir un criterio que nos sirva para buscar el mejor diseño en algún sentido. Su elección dependerá de los intereses que se busquen al realizar el experimento, de la facilidad de cálculo, o de otros aspectos más o menos subjetivos. Damos ahora una primera definición de lo que va a ser una función criterio. Más tarde daremos una definición más completa. Para diferenciar una de otra, a la segunda la llamaremos función criterio convexa.

**Definición 4.1** Diremos que una función:

$$\Phi : \mathcal{M} \longrightarrow \mathbb{R} \cup \{+\infty\}$$

acotada inferiormente es una función criterio si se cumple lo siguiente:

$$\text{var}_{\xi}g \leq \text{var}_{\eta}g \text{ para todo funcional } g \implies \Phi[M(\xi)] \leq \Phi[M(\eta)]$$

Diremos entonces que se trata de un criterio de  $\Phi$ -optimización. Un diseño que minimice  $\Phi[M(\xi)]$  se denominará diseño  $\Phi$ -óptimo.

**Proposición 4.1** Toda función criterio  $\Phi$ , tal y como se ha definido anteriormente es siempre decreciente, en el sentido siguiente:

$$M(\xi) \geq M(\eta) \implies \Phi[M(\xi)] \leq \Phi[M(\eta)]$$

**Demostración:** En efecto, si  $M(\xi) \geq M(\eta)$  entonces por la proposición 2.4 se cumplirá que  $\text{var}_{\xi}g \leq \text{var}_{\eta}g$  para todo funcional  $g$ .

**Observación 4.1** *Puede ocurrir que dos funciones criterio den lugar a un mismo criterio de optimización. Es precisamente lo que demostrará el teorema de equivalencia para algunos criterios que veremos más adelante.*

**Notación 4.1** *Llamaremos  $\mathcal{M}_+ = \{M \in \mathcal{M} : \det M > 0\}$ .  $\mathcal{L}(\mathcal{M})$  es el subespacio vectorial de las matrices simétricas de orden  $m$  generado por el conjunto  $\mathcal{M}$ . Si  $\Phi$  es una función criterio definimos los conjuntos:*

$$\mathcal{M}_\Phi = \{M \in \mathcal{M} : \Phi(M) < \infty\}$$

$$\Xi^* = \{\xi \in \Xi : \xi \text{ es } \Phi\text{-óptimo}\}$$

**Proposición 4.2** *Si  $\Phi$  es una función convexa entonces el conjunto  $\Xi^*$  es convexo.*

**Demostración:** Supongamos que:

$$\Phi[M(\xi_1)] = \Phi[M(\xi_2)] = \min_{\xi \in \Xi} \Phi[M(\xi)]$$

entonces, por ser  $\Phi$  convexa tendremos:

$$\begin{aligned} \Phi\{M[(1-\beta)\xi_1 + \beta\xi_2]\} &= \Phi\{(1-\beta)M(\xi_1) + \beta M(\xi_2)\} \leq \\ &(1-\beta)\Phi[M(\xi_1)] + \beta\Phi[M(\xi_2)] = \min_{\xi \in \Xi} \Phi[M(\xi)] \end{aligned}$$

**Observación 4.2** *En algunas ocasiones no es necesario estimar más que un subconjunto de los parámetros o algunas funciones de ellos, de modo que la función criterio restringe su atención a dichos parámetros. Hablaremos entonces de criterios de optimización parcial o criterios de optimización singulares. Un ejemplo de criterio de optimización parcial viene dado por la función:*

$$\Phi : \mathcal{M} \longrightarrow \mathfrak{R} \cup \{+\infty\} \quad | \quad M(\xi) \longrightarrow \text{var}_{\xi} g$$

*Aunque la función criterio global sea continua, la función criterio parcial correspondiente no siempre es continua en todo  $\mathcal{M}$ . Además el diseño  $\Phi$ -óptimo tiene matriz de información singular en algunos casos. Si los parámetros en los que estamos interesados son los  $s$  primeros entonces la función criterio centrará su atención en las varianzas y covarianzas de estos  $s$  parámetros, es decir en la caja superior izquierda de orden  $s \times s$  de la matriz de información.*

*A partir de los criterios que se estudiarán en el siguiente capítulo pueden definirse criterios parciales de manera natural. No se discutirán en este libro. Una referencia de ellos puede encontrarse, entre otros libros y artículos en [168].*

## 4.2. Funciones criterio y sus propiedades

Para la definición y propiedades de algunas funciones criterio necesitaremos las siguientes definiciones:

**Definición 4.2** Sea  $M_m(\mathbb{R}) \cong \mathbb{R}^{m \times m}$  el conjunto de las matrices cuadradas con términos reales de orden  $m$ . En este espacio vectorial se puede definir el siguiente producto escalar:

$$\langle A, B \rangle = \text{Tr} A^t B$$

de donde la norma de una matriz cuadrada cualquiera de orden  $m$  vendría dada por:

$$\| A \| = (\text{Tr} A^t A)^{1/2}$$

Sabemos que cualquier otra norma definida en este espacio euclídeo es topológicamente equivalente a ésta. En lo sucesivo utilizaremos esta norma.

**Definición 4.3** Sea  $\Phi$  una función definida en un entorno de la matriz  $A$  en el espacio  $M_m(\mathbb{R})$ . Se define el gradiente de  $\Phi$  en la matriz  $A$  como la matriz de componentes:

$$\{\nabla(A)\}_{ij} = \frac{\partial(A)}{\partial A_{ij}}, i, j = 1, \dots, m$$

Damos ya la definición de las funciones criterio más utilizadas:

#### 4.2.1. D-optimización

**Definición 4.4** El criterio de D-optimización viene definido por la función criterio siguiente:

$$\Phi_D[M(\xi)] = \begin{cases} \log \det M^{-1}(\xi) = -\log \det M(\xi) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

**Proposición 4.3** 1.  $\Phi_D$  es continua en  $\mathcal{M}$ .

2. La función  $\Phi_D$  es convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ .

3. En las matrices en que  $\Phi_D$  es finita, también es diferenciable. Además su gradiente es:

$$\nabla[-\log \det M] = -M^{-1}$$

**Demostración:** Puede encontrarse en [168], pg 81.

**Corolario 4.1** Siempre se puede conseguir que el número de puntos en el soporte del diseño D-óptimo esté comprendido entre  $m$  y  $\frac{m(m+1)}{2}$ .

**Demostración:** La cota inferior es consecuencia inmediata de la proposición 2.2. La superior se deduce del hecho de ser la matriz de información una matriz simétrica  $m \times m$  y por su carácter aditivo puede ser expresada como suma de no más de  $\frac{m(m+1)}{2}$  matrices de información de diseños unipuntuales. En efecto, por el teorema de Caratheodory sabemos que si  $\xi$  es un diseño tal que  $M(\xi)$  está en la frontera de  $\mathcal{M}$  entonces existen no más de  $\frac{m(m+1)}{2}$  diseños unipuntuales de modo que  $M(\xi)$  se puede poner como combinación convexa de ellos. En otras palabras, existe un diseño,

$\eta$ , con soporte en no más de  $\frac{m(m+1)}{2}$  puntos tal que  $M(\xi) = M(\eta)$ . Si demostramos que todo diseño D-óptimo,  $\xi^*$ , tiene su matriz asociada en la frontera de  $\mathcal{M}$ , entonces habremos probado lo que queríamos. Supongamos que no es así, y que  $M(\xi^*)$  es un punto interior del conjunto  $\mathcal{M}$ . Existirá entonces un número positivo  $\alpha$  tal que la matriz de información  $(1 + \alpha)M(\xi^*) = M(\mu)$  sigue estando en el conjunto  $\mathcal{M}$ . Ahora bien  $\det M(\mu) = (1 + \alpha) \det M(\xi^*) > \det M(\xi^*)$ , que es contradictorio con el hecho de que  $\xi^*$  es un diseño D-óptimo.

**Observación 4.3** *La gran ventaja de esta función criterio radica en la facilidad de cálculo respecto al resto.*

#### 4.2.2. A-optimización

**Definición 4.5** *El criterio de A-optimización viene definido por la función criterio siguiente:*

$$\Phi_A[M(\xi)] = \sum_{i=1}^m \text{var}_\xi \alpha_i = \begin{cases} \text{tr} M^{-1}(\xi) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

La segunda igualdad es consecuencia del hecho de que:

$$\text{var}_\xi \alpha_i = e_i^t M^{-1}(\xi) e_i$$

si los parámetros son estimables.

**Proposición 4.4**  $\Phi_A$  *tiene las propiedades siguientes:*

1.  $\Phi_A$  es continua en  $\mathcal{M}$ .
2.  $\Phi_A$  es convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ .
3. En las matrices en que  $\Phi_A$  es finita, también es diferenciable. Además su gradiente es:

$$\nabla[\text{tr} M^{-1}(\xi)] = -M^{-2}(\xi)$$

**Demostración:** Puede encontrarse en [168], pg 83..

#### 4.2.3. G-optimización

**Definición 4.6** *El criterio de G-optimización viene definido por la función criterio siguiente:*

$$\Phi_G[M(\xi)] = \sup_{x \in X} \text{var}_\xi g_x, \quad \xi \in \Xi$$

Puesto que:

$$g_x(\theta) = \theta(x), \quad x \in X, \theta \in \Theta$$

entonces:

$$\text{var}_\xi g_x = f^t(x)M^{-1}(\xi)f(x)$$

cuando  $M(\xi)$  es regular. De modo que nuestra función puede escribirse de la forma:

$$\Phi_G[M(\xi)] = \begin{cases} \max_{x \in X} f^t(x)M^{-1}(\xi)f(x) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

**Proposición 4.5** *La función  $\Phi_G$  es:*

1. *continua en  $\mathcal{M}$ .*
2. *convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ .*

**Demostración:** Puede encontrarse en [168], pg 86.

#### 4.2.4. E-optimización

**Definición 4.7** *El criterio de E-optimización viene definido por la función criterio siguiente:*

$$\Phi_E[M(\xi)] = \sup\{\text{var}_\xi \alpha^t c : \|c\| = 1\}, \xi \in \Xi$$

Denotando por  $\lambda_\xi$  el mínimo autovalor de  $M(\xi)$ , la función toma la forma siguiente:

$$\Phi_E[M(\xi)] = \begin{cases} \lambda_\xi^{-1} & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

En efecto,  $\Phi_E[M(\xi)] = \infty$  si, y sólo si existe algún funcional lineal  $\alpha^t c$  que no es estimable para  $\xi$ , es decir, si  $M(\xi)$  es singular. Por otro lado, si  $M(\xi)$  es regular entonces considerando sus autovalores  $\lambda_1, \dots, \lambda_m$  y los correspondientes autovectores  $u_1, \dots, u_m$  se tendrá lo siguiente:

$$M^{-1}(\xi)u_i = M^{-1}(\xi)M(\xi)u_i\lambda_i^{-1} = \lambda_i^{-1}u_i, \text{ para } i = 1, 2, \dots, m$$

de modo que:

$$\sup\{c^t M^{-1}(\xi)c : \|c\| = 1\} \geq u_i^t M^{-1}(\xi)u_i = \lambda_i^{-1}, \text{ para } i = 1, 2, \dots, m$$

Teniendo en cuenta que todo vector  $c$  tal que  $\|c\| = 1$  puede descomponerse de la forma:

$$c = \sum_{i=1}^m a_i u_i, \text{ con } \sum_{i=1}^m c_i^2 = 1$$

de modo que  $c^t M^{-1}(\xi)c = \sum_{i,j} a_i u_i^t M^{-1}(\xi)u_j a_j \leq \lambda_\xi^{-1}$ , ya que  $u_i^t M^{-1}(\xi)u_j$ .

**Proposición 4.6**  *$\Phi_E$  tiene las propiedades siguientes:*

1. *es continua en  $\mathcal{M}$ .*
2. *es convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ .*

**Demostración:** Puede encontrarse en [168], pg 91.

### 4.2.5. c-optimización

Con este criterio el interés se centra en la estimación de combinaciones lineales de los parámetros  $c^t\beta$  con mínima varianza. De este modo se da la siguiente:

**Definición 4.8** *El criterio de c-optimización para un vector  $c$  de dimensión  $m$  viene definido por la función criterio siguiente:*

$$\Phi_c[M(\xi)] = c^t M^{-1}(\xi) c$$

Se toma esta definición puesto que la varianza de  $c^t\beta$  es proporcional a  $c^t M^{-1}(\xi) c$ . La desventaja de los diseños c-óptimos es que son singulares con cierta frecuencia. Elfving, [70], propone un método gráfico para el cálculo del diseño c-óptimo en el caso biparamétrico:

$$y = f(x)\alpha + g(x)\beta + \varepsilon(x), \sigma^2(x) = \sigma^2, x \in X$$

Suponemos que se toma un diseño exacto cualquiera,  $\xi$ , concentrado en los puntos  $x_1, \dots, x_n$ , con pesos  $p_i = N_i/N$ . Es decir, se toman  $N_i$  observaciones en cada punto  $x_i$ , siendo  $N$  el número total de observaciones realizadas. Tenemos así  $n$  puntos del plano real definidos de la forma  $X_1 = (f(x_1), g(x_1)), \dots, X_n = (f(x_n), g(x_n))$ . La media de las observaciones en cada punto  $x_i$  tiene la siguiente ecuación de regresión:

$$\bar{y}_i = f(x_i)\alpha + g(x_i)\beta + \frac{\bar{\varepsilon}_i}{\sqrt{p_i}}$$

donde  $\bar{\varepsilon}_i$  tiene varianza  $\sigma^2/N_i$ . Por simplicidad de notación supondremos que  $\sigma^2/N_i = 1$ , que no significa ninguna merma en la generalidad de los resultados. De este modo  $\text{var}(\frac{\bar{\varepsilon}_i}{\sqrt{p_i}}) = 1$ . Sea ahora  $t = \sum_{i=1}^n a_i y_i$  un estimador lineal de  $c^t\beta$ . Será centrado cuando:

$$E(t) = \sum_{i=1}^n a_i [f(x_i)\alpha + g(x_i)\beta] = c_1\alpha + c_2\beta$$

es decir, se ha de satisfacer la ecuación vectorial:

$$\sum_{i=1}^n a_i X_i = c \quad (4.1)$$

Además para este estimador la varianza se hace mínima cuando se toman los pesos  $\xi_a(x_i) = |a_i|/k_a$ , donde  $k_a = \sum_{i=1}^n |a_i|$ :

$$\min_{\xi} \text{var}\left(\sum_{i=1}^n a_i y_i\right) = \min_{\xi} \sum_{i=1}^n \frac{a_i^2}{\xi(x_i)} = k_a^2$$

Nos interesa ahora tomar los coeficientes  $a_i$  que hagan mínima la varianza anterior bajo la condición (4.1), que podemos escribir de la forma:

$$c = k_a \sum_{i=1}^n \xi_a(x_i) \text{sgn}(a_i) X_i = k_a c_a$$

donde hemos denotado por  $c_a$  el correspondiente vector. El factor  $k_a$  es un número positivo y  $c_a$  es un vector con la misma dirección que  $c$ . Podemos representar esta situación en la Figura 4.1:

Figura 4.1: c-optimización

Los pesos  $\xi_a(x_i)$  son no negativos y de suma uno, por tanto el extremo pertenece al cierre convexo del conjunto  $\{\pm X_1, \dots, \pm X_n\}$ . Es claro que la varianza alcanzará su mínimo cuando el extremo del vector  $c_a$  coincida con la intersección  $A^*$  del vector  $c$ , o su prolongación, con el polígono de la Figura 4.1. Si este punto está, por ejemplo, entre los vértices  $X_1$  y  $X_2$  entonces el diseño c-óptimo tiene como soporte los puntos  $x_1$  y  $x_2$  con pesos proporcionales a las distancias  $A^*X_1$  y  $A^*X_2$  respectivamente. Además, el valor de la varianza viene dado por  $[(OA)/(OA^*)]^2$ . Este método proporciona un instrumento rápido para el cálculo del diseño c-óptimo, y es generalizable a diseños aproximados.

#### 4.2.6. L-optimización

La siguiente definición es debida a Atwood ([17], pg 1125):

**Definición 4.9** *El criterio de L-optimización viene definido por la función criterio siguiente:*

$$\Phi_L[M(\xi)] = \begin{cases} \text{Tr}WM^{-1}(\xi) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

donde  $W$  es una matriz definida positiva de orden  $m$ .

**Proposición 4.7**  $\Phi_L$  tiene las propiedades siguientes:

1.  $\Phi_L$  es continua en  $\mathcal{M}$ .
2.  $\Phi_L$  es convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ .
3. En las matrices en que  $\Phi_L$  es finita, también es diferenciable. Además su gradiente es:

$$\nabla[\text{Tr}WM^{-1}(\xi)] = -M^{-1}(\xi)WM^{-1}(\xi)$$

**Demostración:** El segundo apartado se demostrará más tarde. La demostración a los otros dos puede encontrarse en [168], pg 93.

#### 4.2.7. $\Phi_p$ -optimización

**Definición 4.10** El criterio de  $\Phi_p$ -optimización es el asociado a la siguiente función criterio:

$$\Phi_p[M(\xi)] = \begin{cases} [m^{-1} \text{tr} M^{-p}(\xi)]^{1/p} & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

siendo  $p$  un número real positivo.

Si denotamos por  $\lambda_i \xi$ ,  $i = 1, \dots, m$  los autovalores de  $M(\xi)$  entonces esta función puede expresarse de la forma siguiente:

$$\Phi_p[M(\xi)] = \begin{cases} [m^{-1} \sum_{i=1}^m \lambda_{i,\xi}^{-p}]^{1/p} & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

#### 4.2.8. $L_p$ -optimización

Este caso viene a generalizar gran parte de los criterios de optimización más usuales:

**Definición 4.11** El criterio de optimización  $L_p$  es el asociado a la siguiente función criterio:

$$\Phi_{L_p}[M(\xi)] = \begin{cases} [m^{-1} \text{tr}(HM^{-1}(\xi)H^t)^p]^{1/p} & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

siendo  $p$  un número real positivo y  $H$  una matriz regular de orden  $m$ .

**Proposición 4.8** Si  $p$  es un entero positivo, entonces  $\Phi_{L_p}$  tiene las propiedades siguientes:

1.  $\Phi_{L_p}$  es continua en  $\mathcal{M}$ .
2.  $\Phi_{L_p}$  es convexa en  $\mathcal{M}$  y estrictamente convexa en  $\mathcal{M}_+$ .
3. En las matrices en que  $\Phi_{L_p}$  es finita, también es diferenciable. Además su gradiente,

$$\nabla \Phi_{L_p}[M(\xi)] = \frac{-1}{pm^{1/p}} \{m^1 \text{tr}[HM^{-1}(\xi)H^t]^p\}^{\frac{1}{p-1}} \sum_{k=0}^{p-1} M^{k-p}(\xi)(H^t)^p H^p M^{-k-1}(\xi)$$

**Demostración:** El segundo apartado se demostrará más tarde. La demostración a los otros dos puede encontrarse en [168], pg 83.

#### 4.2.9. I-optimización

Teniendo en cuenta que  $f^t(x_0)M^{-1}(\xi)f(x_0)$  es la varianza de la predicción de la respuesta en el punto  $x$ , este criterio se interesará en minimizar el valor esperado de dichas varianzas sobre el conjunto  $X$ . Suponiendo que  $\mu$  es una medida de probabilidad sobre  $X$  se define la función criterio como:

$$\Phi_V[M(\xi)] = \begin{cases} \int_X f^t(x)M^{-1}(\xi)f(x)d\mu & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

#### 4.2.10. MV-optimización

Dicho criterio buscará minimizar el máximo de las varianzas de los estimadores de los parámetros. De este modo se centra la atención en las varianzas de los estimadores de todos los parámetros a la vez, sin tener en cuenta las covarianzas. La función criterio será entonces:

$$\Phi_{MV}[M(\xi)] = \begin{cases} \max_i \text{var}_\xi \alpha_i & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

es decir, se tratará de minimizar el máximo de los elementos de la diagonal de la matriz inversa de información:

$$\Phi_{MV}[M(\xi)] = \begin{cases} \max_i \{M^{-1}(\xi)\} & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

#### 4.2.11. Criterio minimax

Denotando por  $|\cdot|$  una norma cualquiera en el espacio  $\mathfrak{R}^m$ , se define la función criterio:

$$\Phi_{|\cdot|}[M(\xi)] = \begin{cases} \max\{c^t M^{-1}(\xi) c : c \in \mathfrak{R}^m, |c| = 1\} & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

El siguiente teorema muestra la relación entre algunos criterios:

**Teorema 4.1** 1. Si en la  $L_p$ -optimización hacemos  $H = I$  y hacemos tender  $p$  a infinito obtenemos  $E$ -optimización.

2. Si se toma  $W = m^{-1} H^t H$  y  $p = 1$  la  $L_p$ -optimización se transforma en  $L$ -optimización.

3. Haciendo  $H = I$  y  $p$  que tienda a cero la  $L_p$ -optimización se convierte en  $D$ -optimización.

4. Si en la  $L$ -optimización hacemos  $W = I$  coincidirá con  $A$ -optimización.

5. La  $\Phi_p$ -optimización es un caso particular de  $L_p$ -optimización tomando  $H = I$ .

6. Llamando  $W = \int_X f^t(x) f(x) d\mu$  en  $L$ -optimización se tiene  $I$ -optimización.

7. Si en el criterio minimax se utiliza la norma  $|\cdot|_1$  entonces obtenemos el criterio  $MV$ , utilizando la norma  $|\cdot|_2$  se obtiene el criterio  $E$  y utilizando la siguiente norma se obtiene el criterio  $G$ :

$$|c|^{\mathcal{R}} = \inf\{\alpha \geq 0 : c \in \alpha \mathcal{R}\}$$

donde  $\mathcal{R}$  es el cierre convexo del conjunto:

$$\{f(x) : x \in X\} \cup \{-f(x) : x \in X\}$$

**Demostración:** Para demostrar los apartados segundo, cuarto y quinto bastará hacer la sustitución indicada. Para probar el primer apartado denotemos por  $\lambda_1, \dots, \lambda_m$  los autovalores de  $M(\xi)$  y sea  $\lambda_{i_0}$  el mínimo de los autovalores. Entonces:

$$\lim_{p \rightarrow \infty} [\text{tr} M^{-p}(\xi)]^{1/p} = \lim_{p \rightarrow \infty} \left[ \sum_{i=1}^m \lambda_i^{-p} \right]^{1/p} = \lim_{p \rightarrow \infty} \lambda_{i_0}^{-1} \left[ 1 + \sum_{i \neq i_0} \left( \frac{\lambda_i}{\lambda_{i_0}} \right)^{-p} \right]^{1/p} = \lambda_{i_0}^{-1}$$

Para demostrar el tercer apartado hacemos  $H = I$  de modo que:

$$\lim_{p \rightarrow 0} \log \left[ \frac{1}{m} \operatorname{tr} M^{-p}(\xi) \right]^{1/p} = \lim_{p \rightarrow 0} \frac{\log \frac{1}{m} \sum_{i=1}^m \lambda_i^{-p}}{p} = \frac{-1}{m} \log \det M(\xi)$$

Veamos ahora que se cumple el sexto resultado. En efecto, tomando en L-optimización  $W = \int_X f^t(x) f(x) d\mu$  se tendrá:

$$\begin{aligned} \operatorname{tr} W M^{-1}(\xi) &= \operatorname{tr} \int_X f^t(x) f(x) d\mu M^{-1}(\xi) = \operatorname{tr} \int_X f^t(x) f(x) M^{-1}(\xi) d\mu = \\ &= \operatorname{tr} \int_X f^t(x) M^{-1}(\xi) f(x) d\mu = \int_X f^t(x) M^{-1}(\xi) f(x) d\mu \end{aligned}$$

Para probar el séptimo apartado vemos que:

$$\Phi_{|\cdot|_1}[M(\xi)] = \max\{c^t M^{-1}(\xi) c : c \in \mathfrak{R}^m, |c|_1 = 1\} = \max_i e_i^t M^{-1}(\xi) e_i$$

siendo  $e_1, \dots, e_m$  la base canónica. La segunda igualdad es debida al hecho de que  $c^t M^{-1}(\xi) c$  es una función convexa en  $c$  y por tanto el máximo se alcanzará en los vértices del conjunto  $\{c \in \mathfrak{R}^m : |c|_1 = 1\}$ , que vienen dados por los vectores de la base canónica. Por otra parte,

$$\Phi_{|\cdot|_{\mathcal{R}}}[M(\xi)] = \max\{c^t M^{-1}(\xi) c : c \in \partial \mathcal{R}\} = \max_{x \in X} f^t(x) M^{-1}(\xi) f(x)$$

Para justificar la primera y segunda igualdades bastará observar que la norma  $|\cdot|_{\mathcal{R}}$  caracteriza el conjunto  $\mathcal{R}$  como la bola unidad y el borde de la misma será precisamente el conjunto:

$$\{f(x) : x \in X\} \cup \{-f(x) : x \in X\}$$

Otros criterios con mayor o menor grado de generalidad se han definido en la literatura. Los que hemos dado son un ejemplo de los más utilizados.

Cuando hablemos de los procesos iterativos que conducen al cálculo del diseño óptimo tendremos que calcular frecuentemente la matriz de información inversa. Puesto que en cada paso utilizamos el diseño construido en el paso anterior, parece interesante obtener una expresión explícita de esta inversa en función de la inversa calculada en el paso anterior. No cabe duda de que esto reduce los cálculos considerablemente y hace el problema fácilmente programable. A este fin va dirigida la siguiente proposición:

**Proposición 4.9** Sea  $\eta = (1 - \beta)\xi + \beta\xi_x$ , donde  $\beta \in (0, 1)$  y  $\xi \in \Xi$ . Suponemos además que  $\det M(\xi) \neq 0$ . Entonces  $\det M(\eta) \neq 0$ , y:

$$\begin{aligned} M^{-1}(\eta) &= \frac{1}{1 - \beta} \left\{ M^{-1}(\xi) - \frac{\beta M^{-1}(\xi) f(x) f^t(x) M^{-1}(\xi)}{1 - \beta + \beta f^t(x) M^{-1}(\xi) f(x)} \right\} \\ \det M^{-1}(\eta) &= \det M^{-1}(\xi) (1 - \beta)^{-m} \left[ 1 - \frac{\beta f^t(x) M^{-1}(\xi) f(x)}{1 - \beta + \beta f^t(x) M^{-1}(\xi) f(x)} \right] \end{aligned}$$

**Demostración:** Puede encontrarse en [168], pg. 134.

#### Ejemplo 4.1

Siguiendo con el ejemplo 2.3 calcularemos el valor de cada una de estas funciones criterio en los tres diseños que se han utilizado.

$$\begin{aligned}\Phi_D[M(\xi_0)] &= \log \det M^{-1}(\xi_0) = -\log \frac{9}{25\sigma^4} = 1,021165 + 4 \log \sigma \\ \Phi_D[M(\xi_1)] &= \log \det M^{-1}(\xi_1) = -\log \frac{1}{4\sigma^4} = 1,3863 + 4 \log \sigma \\ \Phi_D[M(\xi_2)] &= \log \det M^{-1}(\xi_2) = -\log \frac{1}{4\sigma^4} = 4 \log \sigma\end{aligned}$$

de modo que de los tres diseños  $\xi_2$  es el mejor de acuerdo con este criterio.

$$\begin{aligned}\Phi_A[M(\xi_0)] &= \text{tr}M^{-1}(\xi_0) = \frac{10\sigma^2}{3} \\ \Phi_A[M(\xi_1)] &= \text{tr}M^{-1}(\xi_1) = 4\sigma^2 \\ \Phi_A[M(\xi_2)] &= \text{tr}M^{-1}(\xi_2) = 2\sigma^2\end{aligned}$$

y  $\xi_2$  sigue siendo el mejor de acuerdo con criterio de A-optimización. Lo mismo va a ocurrir con los demás criterios que veremos:

$$\begin{aligned}\Phi_G[M(\xi_0)] &= \max_{x \in X} f^t(x)M^{-1}(\xi_0)f(x) = \frac{10\sigma^2}{3} \\ \Phi_G[M(\xi_1)] &= \max_{x \in X} f^t(x)M^{-1}(\xi_1)f(x) = 4\sigma^2 \\ \Phi_G[M(\xi_2)] &= \max_{x \in X} f^t(x)M^{-1}(\xi_2)f(x) = 2\sigma^2\end{aligned}$$

En este caso coincide con los valores de la función criterio de A-optimización. Siendo  $\lambda_\xi$  el mínimo autovalor de  $M(\xi)$  se tiene:

$$\begin{aligned}\Phi_E[M(\xi_0)] &= \lambda_\xi^{-1} = \frac{5\sigma^2}{3} \\ \Phi_E[M(\xi_1)] &= \lambda_\xi^{-1} = 2\sigma^2 \\ \Phi_E[M(\xi_2)] &= \lambda_\xi^{-1} = \sigma^2\end{aligned}$$

Puesto que las tres matrices son simétricas, los valores de la función criterio E coinciden con los valores de la función criterio MV. Sea ahora  $W$  una matriz simétrica definida positiva, entonces:

$$\begin{aligned}\Phi_L[M(\xi_0)] &= \text{tr}[WM^{-1}(\xi_0)] = \frac{5\sigma^2}{3}TrW \\ \Phi_L[M(\xi_1)] &= \text{tr}[WM^{-1}(\xi_1)] = 2\sigma^2TrW \\ \Phi_L[M(\xi_2)] &= \text{tr}[WM^{-1}(\xi_2)] = \sigma^2TrW\end{aligned}$$

Si  $p > 0$  entonces:

$$\begin{aligned}\Phi_p[M(\xi_0)] &= [m^{-1}TrM^{-p}(\xi_0)]^{1/p} = \left\{ \frac{1}{2} \left[ \left( \frac{5\sigma^2}{3} \right)^p + \left( \frac{5\sigma^2}{3} \right)^p \right] \right\}^{1/p} = \frac{5\sigma^2}{3} \\ \Phi_p[M(\xi_1)] &= [m^{-1}TrM^{-p}(\xi_1)]^{1/p} = 2\sigma^2 \\ \Phi_p[M(\xi_2)] &= [m^{-1}TrM^{-p}(\xi_2)]^{1/p} = \sigma^2\end{aligned}$$

Por último si  $H$  es una matriz regular y  $p > 0$  se tendrá:

$$\begin{aligned}\Phi_{L_p}[M(\xi_0)] &= [m^{-1}\text{tr}(HM^{-1}(\xi_0)H^t)^p]^{1/p} = \frac{5\sigma^2}{3} \frac{1}{2^{1/p}} [\text{tr}(HH^t)^p]^{1/p} \\ \Phi_{L_p}[M(\xi_1)] &= [m^{-1}\text{tr}(HM^{-1}(\xi_1)H^t)^p]^{1/p} = 2\sigma^2 \frac{1}{2^{1/p}} [\text{tr}(HH^t)^p]^{1/p} \\ \Phi_{L_p}[M(\xi_2)] &= [m^{-1}\text{tr}(HM^{-1}(\xi_2)H^t)^p]^{1/p} = \sigma^2 \frac{1}{2^{1/p}} [\text{tr}(HH^t)^p]^{1/p}\end{aligned}$$

Para la I-optimización utilizando la medida de contar sobre  $X$  tendremos:

$$\begin{aligned}\Phi_V[M(\xi_0)] &= \sum_{i=0}^4 f^t(x_i)M^{-1}(\xi_0)f(x_i) = \\ &= \frac{5\sigma^2}{3} \left[ (1,0) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (0,1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \right. \\ & \left. (1,1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (1,-1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right] = 10\sigma^2 \\ \Phi_V[M(\xi_1)] &= 12\sigma^2 \\ \Phi_V[M(\xi_2)] &= 6\sigma^2\end{aligned}$$

Los casos más interesantes del criterio minimax (E-, MV- y G-optimización) ya han sido tratados. En el caso de c-optimización podemos tomar un vector cualquiera  $c^t = (a, b)$ , de modo que la función criterio tomará los valores:

$$\begin{aligned}\Phi_c[M(\xi_0)] &= (a, b)M^{-1}(\xi_0) \begin{pmatrix} a \\ b \end{pmatrix} = \frac{5\sigma^2}{3}(a, b) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \frac{5\sigma^2}{3}(a + b) \\ \Phi_c[M(\xi_1)] &= \frac{6\sigma^2}{5}(a + b) \\ \Phi_c[M(\xi_2)] &= \frac{10\sigma^2}{9}(a + b)\end{aligned}$$

## Capítulo 5

# El teorema de equivalencia

Kiefer y Wolfowitz demostraron en 1960 ([129]) la equivalencia de los criterios D-óptimo y G-óptimo. Esto quiere decir que utilizando ambos criterios siempre llegaremos al mismo diseño óptimo. En concreto:

**Teorema 5.1** *Suponiendo que  $\sigma(x)$  es constante para todo  $x$  de  $X$ , un diseño  $\xi^*$  es D-óptimo si, y sólo si es G-óptimo. Es decir son equivalentes:*

1.  $\det M(\xi^*) = \max\{\det M(\xi) : \xi \in \Xi\}$
2.  $\max_{x \in X} f^t(x)M^{-1}(\xi^*)f(x) = \min_{\xi \in \Xi} \max_{x \in X} f^t(x)M^{-1}(\xi)f(x)$

Además, la última expresión es igual a  $m$ .

**Observación 5.1** 1. *Si  $\sigma(x)$  no es constante, el teorema de equivalencia se cumple también haciendo una pequeña precisión. En este caso puede demostrarse que  $\xi^*$  es D-óptimo si, y sólo si*

$$\max_{x \in X} \sigma^{-2}(x) f^t(x) M^{-1}(\xi^*) f(x) = \min_{\xi \in \Xi} \max_{x \in X} \sigma^{-2}(x) f^t(x) M^{-1}(\xi) f(x)$$

Siendo además la última expresión igual a  $m$ .

2. *El teorema de equivalencia no se cumple para diseños exactos: Se puede demostrar que el diseño concentrado en  $-1, 0$  y  $1$  con pesos  $1/4, 1/2$  y  $1/4$  es D-óptimo exacto de tamaño 4 con  $\det M(\xi^*) = 0,125$ . Y sin embargo el diseño concentrado en  $-1, -\sqrt{\sqrt{5}-2}, \sqrt{\sqrt{5}-2}$  y  $1$  con peso  $1/4$  en cada uno es G-óptimo exacto de tamaño 4 y  $\det M(\xi^*) = 0,0902$ .*

**Observación 5.2** 1. *El volumen del elipsoide  $m$ -dimensional:*

$$\mathcal{E} = \{z \in \mathfrak{R}^m : z^t M(\xi) z \leq c^2\}$$

es igual a  $c^m V_m [\det M^{-1}(\xi)]^{1/2}$ , donde  $V_m$  es el volumen de la esfera  $m$ -dimensional de radio unidad. En efecto, por ser  $M(\xi)$  simétrica existirá una matriz regular  $U$  tal que  $U^t M(\xi) U = I$ .

Llamando  $w = \frac{U^{-1}z}{c}$  tendremos:

$$\text{Vol}(\mathcal{E}) = \int_{\mathcal{E}} dz = \int_{\mathcal{S}} c^m |\det U| dw = c^m V_m |\det U|$$

donde haciendo el cambio de variable anterior la región  $E$  se transforma en la región:  $\mathcal{S} = \{w : c^2 w^t U^t M(\xi) U w \leq c^2\} = \{w : w^t w \leq c^2\}$ , que es la esfera  $m$ -s. Pero  $(\det U)^2 \det M(\xi) = 1$ , con lo que queda probado el resultado.

2. Si se supone que  $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_m)$  se distribuye de acuerdo a una distribución normal multivariante de modo que la media tiene como componentes los verdaderos valores de los parámetros  $\alpha$  y la matriz de covarianzas es  $M^{-1}(\xi)$ , entonces la función de densidad de este vector aleatorio será:

$$f(\hat{\alpha} | \alpha) = \frac{[\det M(\xi)]^{1/2}}{(2\pi)^{m/2}} \exp\left\{-\frac{1}{2}(\hat{\alpha} - \alpha)^t M(\xi)(\hat{\alpha} - \alpha)\right\}$$

Consideremos entonces el elipsoide siguiente:

$$\mathcal{E} = \{\hat{\alpha} \in \mathbb{R}^m : (\hat{\alpha} - \alpha)^t M(\xi)(\hat{\alpha} - \alpha) \leq c^2\},$$

donde  $c$  es un número fijo. Además la variable  $v = U^{-1}(\hat{\alpha} - \alpha)$  sigue una distribución normal de media  $0$  y vector de covarianzas la identidad. De este modo  $v^t v$  sigue una distribución  $\chi_m^2$ . Por tanto ahora:

$$\int_{\mathcal{E}} f(\hat{\alpha} | \alpha) d\hat{\alpha} = \int_{\{v^t v \leq c^2\}} (2\pi)^{-m/2} \exp\left[-\frac{1}{2}v^t v\right] dv = P(\chi_m^2 \leq c^2).$$

Se deduce, por tanto, que el elipsoide:

$$\{z \in \mathbb{R}^m : (z - \hat{\alpha})^t M(\xi)(z - \hat{\alpha}) \leq c^2\}$$

contiene la verdadera media con probabilidad  $P(\chi_m^2 \leq c^2)$ , es decir, es el elipsoide de confianza de la media  $\alpha$  con un nivel de confianza que lógicamente depende de  $c$ .

3. Por otro lado el elipsoide:

$$\mathcal{E}_\xi = \{z \in \mathbb{R}^m : z^t M^{-1}(\xi) z \leq m\}$$

contendrá los vectores  $f(x), x \in X$ , es decir, contendrá al conjunto  $f(X)$  solamente si  $\xi$  es un diseño  $D$ -óptimo. Esto es debido al teorema de equivalencia, que asegura lo siguiente:

$$\min_{\xi \in \Xi} \max_{x \in X} f^t(x) M^{-1}(\xi) f(x) = m$$

de modo que solamente si  $\xi$  es un diseño  $D$ -óptimo se cumple:

$$f^t(x) M^{-1}(x) f(x) \leq m, x \in X$$

Además como se vió arriba el volumen de  $\mathcal{E}_\xi$  es proporcional a  $[\det M(\xi)]^{1/2}$ . De modo que el diseño  $D$ -óptimo será aquel que haga máximo el volumen del elipsoide  $\mathcal{E}_\xi$  conteniendo el conjunto  $f(X)$ . El problema dual sería encontrar la matriz  $M$  que minimiza el elipsoide:

$$\mathcal{E}_\xi^* = \{z \in \mathbb{R}^m : z^t M z \leq m\}$$

sujeto a la restricción  $f(X) \subset \mathcal{E}_\xi^*$ . El principio fuerte de Lagrange asegura que existe un valor extremo común para ambos problemas. Uno de los inconvenientes de este criterio es que a veces el volumen del elipsoide puede ser minimizado haciéndose estrecho y largo. Eso significa que hay un funcional lineal de los parámetros, que es estimado con una gran varianza. No obstante no es sencillo llevar a la práctica este resultado para el cálculo del  $D$ -óptimo. Silvey, [187], proporciona un método para su utilización.

En 1973 Whittle, [214], y en 1974 Kiefer J., [124], generalizan el teorema de equivalencia para funciones criterio más generales. Con el objeto de enunciar estos resultados damos una definición más rigurosa de función criterio. Para ello nos fijaremos en las propiedades que verificaban las funciones criterio vistas anteriormente. Observamos que todas las funciones criterio vistas son convexas y continuas. Con el objeto de salvaguardar estas propiedades daremos la siguiente definición:

**Definición 5.1** Diremos que una función real  $\Phi$  en las condiciones de una función criterio, según la definición dada más arriba, es una función criterio convexa si cumple las propiedades:

1. Existe  $U_\Phi$  abierto de  $\mathcal{L}(\mathcal{M})$  tal que  $U_\Phi \supset \mathcal{M}_+$  y  $\Phi$  está definida, es finita y convexa en  $U_\Phi$ .
2. Si:

$$M_n \in \mathcal{M}_+, n = 1, 2, \dots \text{ y } \lim_{n \rightarrow \infty} M_n = M \in \mathcal{M} - \mathcal{M}_+$$

entonces:

$$\lim_{n \rightarrow \infty} \Phi(M_n) = \infty$$

**Observación 5.3** No es necesario que  $U_\Phi$  sea convexo. Entenderemos entonces que  $\Phi$  debe ser convexa en los subconjuntos convexos de  $U_\Phi$ .

**Proposición 5.1** Si  $U$  es un abierto de  $\mathcal{L}(\mathcal{M})$  y  $\Phi$  es convexa y finita en  $U$ , entonces  $\Phi$  es continua en  $U$ .

**Demostración:** Una demostración de este resultado puede encontrarse en [110].

**Proposición 5.2** Cualquier función criterio convexa, con la definición dada, es continua en  $\mathcal{M}$ .

**Demostración:** La proposición anterior implica la continuidad en  $U_\Phi$ , y por tanto en  $\mathcal{M}_+$ . Por otra parte la segunda condición de la definición asegura la continuidad en  $\mathcal{M} - \mathcal{M}_+$ .

Todavía necesitamos dar una definición de derivada direccional debida a Frechét:

**Definición 5.2** Dada una función real  $\Phi$ , convexa y definida en un subconjunto convexo de un espacio euclídeo, y dados dos puntos de ese conjunto  $x$  y  $v$ , se define la derivada direccional de  $\Phi$  en el punto  $x$  y en la dirección  $v$  como:

$$\partial\Phi(x, v) = \lim_{\beta \rightarrow 0^+} \frac{\Phi[(1 - \beta)x + \beta v] - \Phi(x)}{\beta}$$

**Observación 5.4** *La derivada direccional existe siempre gracias a la convexidad de  $\Phi$ . En efecto, demostraremos que la función:*

$$\varphi : (0, 1) \longrightarrow \Re \mid \beta \longrightarrow \frac{\Phi[(1 - \beta)x + \beta v] - \Phi(x)}{\beta}$$

*es creciente en  $(0, 1)$ . Por tanto el límite siguiente siempre existe, siendo finito o infinito negativo:*

$$\lim_{\beta \rightarrow 0^+} \frac{\Phi[(1 - \beta)x + \beta v] - \Phi(x)}{\beta}$$

*Veamos que  $\varphi$  es una función creciente. Para ello tomamos  $0 < \beta_1 < \beta_2 < 1$  y calculamos:*

$$\begin{aligned} \Phi[(1 - \beta_1)x + \beta_1 v] - \Phi(x) &= \Phi\left\{\frac{\beta_1}{\beta_2}[(1 - \beta_2)x + \beta_2 v] + \left(1 - \frac{\beta_1}{\beta_2}\right)x\right\} - \Phi(x) \leq \\ &= \frac{\beta_1}{\beta_2}\Phi[(1 - \beta_2)x + \beta_2 v] + \left(1 - \frac{\beta_1}{\beta_2}\right)\Phi(x) - \Phi(x) = \\ &= \frac{\beta_1}{\beta_2}\{\Phi[(1 - \beta_2)x + \beta_2 v] - \Phi(x)\} \end{aligned}$$

*Pueden darse otras definiciones de derivada direccional, pero nos interesará la anterior debido al uso de funciones convexas. Si existe el gradiente de  $\Phi$  entonces por su definición, podemos escribir:*

$$\partial\Phi(x, v) = \text{tr}\{[\nabla\Phi(x)](v - x)\} = \langle \nabla\Phi(x), v - x \rangle$$

*Recíprocamente, se puede definir el gradiente como aquella matriz que encaja en la expresión anterior.*

**Teorema 5.2 (teorema general de equivalencia)** *Sea  $\Phi$  una función criterio convexa y  $\xi^*$  un diseño tal que:*

$$\partial\Phi[M(\xi^*), M(\xi)] > -\infty, \xi \in \Xi$$

*Entonces son equivalentes:*

1.  $\xi^*$  es  $\Phi$ -óptimo.
2.  $\xi^*$  es  $\Phi$ -óptimo local, es decir, para cada diseño  $\xi$  la función:

$$[0, 1) \longrightarrow \Re \mid \beta \longrightarrow \Phi[(1 - \beta)M(\xi^*) + \beta M(\xi)]$$

*tiene un mínimo local en  $\beta = 0$ .*

3.  $\partial\Phi[M(\xi^*), M(\xi)] \geq 0, \xi \in \Xi$ .

Obsérvese que este teorema aporta un criterio general (tercer apartado) para contrastar si un diseño dado es o no  $\Phi$ -óptimo, tanto si la función criterio es diferenciable como si no lo es. Cuando la función  $\Phi$  sea diferenciable, se puede dar el siguiente teorema:

**Teorema 5.3** *Si  $\Phi$  es diferenciable en un entorno de  $M(\xi^*)$ , entonces son equivalentes:*

1.  $\xi^*$  es  $\Phi$ -óptimo.
2.  $f^t(x) \nabla \Phi[M(\xi^*)]f(x) \geq \text{Tr}M(\xi^*) \nabla \Phi[M(\xi^*)], x \in X$ .
3.  $\min_{x \in X} f^t(x) \nabla \Phi[M(\xi^*)]f(x) = \sum_{x \in X} f^t(x) \nabla \Phi[M(\xi^*)]f(x)\xi^*(x)$

## Capítulo 6

# Construcción del diseño óptimo

### 6.1. Aproximación al diseño óptimo

En la mayoría de los casos que nos encontramos en la práctica, no va a ser posible encontrar exactamente el diseño óptimo bajo el aspecto de un determinado criterio. Procederemos entonces según un algoritmo convergente encaminado a la obtención de un diseño aproximado al óptimo. Pero, ¿cómo saber si un diseño obtenido en un paso del algoritmo está o no suficientemente próximo al diseño óptimo?. A resolver esta cuestión van encaminados los siguientes resultados debidos a Silvey, [184].

**Proposición 6.1** *Sea  $\Phi$  una función criterio convexa. Si  $\mu$  es un diseño tal que  $\Phi[M(\mu)] < \infty$  y para un cierto  $\delta > 0$  se tiene:*

$$\partial\Phi[M(\mu), M] \geq -\delta, M \in \mathcal{M} \quad (6.1)$$

entonces:

$$\Phi[M(\mu)] \leq \inf\{\Phi[M(\xi)] : \xi \in \Xi\} + \delta$$

Es decir, si  $\xi^*$  es  $\Phi$ -ptimo entonces:

$$\Phi[M(\mu)] - \Phi[M(\xi^*)] \leq \min_{M \in \mathcal{M}} \partial\Phi[M(\mu), M]$$

Si además existe el gradiente  $\nabla\Phi[M(\mu)]$  entonces (6.1) es equivalente a:

$$f^t(x) \nabla \Phi[M(\mu)] f(x) \geq \text{tr} \nabla \Phi[M(\mu)] M(\mu) - \delta, x \in X.$$

En 1969 Atwood, [15], dio otra acotación del error cometido al aceptar como D-ptimo un diseño dado:

**Proposición 6.2** *Si  $\det M(\mu) \neq 0$ , entonces:*

$$\left(\frac{\det M(\mu)}{\max_{\xi \in \Xi} \det M(\xi)}\right)^{1/m} \geq \frac{m}{f^t(x)M^{-1}(\mu)f(x)}$$

y tomando logaritmos:

$$\frac{1}{m}(\Phi_D[M(\mu)] - \inf\{\Phi_D[M(\xi)] : \xi \in \Xi\}) \leq \log \frac{\max_{x \in X} f^t(x)M^{-1}(\mu)f(x)}{m}$$

## 6.2. Procesos iterativos para el cálculo del diseño D-óptimo

Los algoritmos que se dan a continuación se basarán en el gradiente. Se define a partir del gradiente de la función criterio de D-optimización la siguiente función:

$$d(x, \xi) = \sigma^{-2}(x) f^t(x) M^{-1}(\xi) f(x)$$

Para detener el algoritmo puede utilizarse la Proposición 6.2, que puede expresarse de la forma:

$$\Phi_D[M(\xi)] - \Phi_D[M(\xi^*)] \leq m \log[\max_{x \in X} d(x, \xi)/m]$$

siendo  $\xi$  un diseño tal que  $\det M(\xi) \neq 0$  y  $\xi^*$  un diseño D-óptimo. Damos a continuación algunos de los algoritmos más utilizados.

### 6.2.1. Algoritmo de paso armónico

Puede encontrarse en [168], Sección V.2.

1. Se elige un diseño inicial  $\xi_0$  tal que,  $\det M(\xi_0) \neq 0$ .
2. Dado un  $\xi_n$  se define un nuevo diseño:

$$\xi_{n+1} = \left(1 - \frac{1}{n+2}\right) \xi_n + \frac{1}{n+2} \xi_{x_n}$$

donde  $x_n$  es el punto donde se alcanza el máximo de la función:

$$X \longrightarrow \Re \mid x \longrightarrow d(x, \xi_n)$$

3. El proceso se detiene de acuerdo con la Proposición 6.2.

### 6.2.2. Algoritmo ascendente

Ahora el paso,  $\beta$ , se mejorará en cada etapa del algoritmo.

1. Se elige un diseño inicial  $\xi_0$  tal que,  $\det M(\xi_0) \neq 0$ .
2. Dado un  $\xi_n$  se define un nuevo diseño:

$$\xi_{n+1} = (1 - \beta_n) \xi_n + \beta_n \xi_{x_n}$$

donde  $x_n$  es el punto donde se alcanza el máximo de la función:

$$X \longrightarrow \Re \mid x \longrightarrow d(x, \xi_n)$$

y  $\beta_n$  es el punto donde se alcanza el máximo de:

$$[0, 1] \longrightarrow \Re \mid \beta \longrightarrow \det M[(1 - \beta_n) \xi_n + \beta_n \xi_{x_n}]$$

3. El proceso se detiene de acuerdo con la Proposición 6.2.

**Proposición 6.3** ([74], Sección 2.5) , *En el proceso anterior se puede calcular  $\beta_n$  a partir de la fórmula:*

$$\beta_n = \frac{d(x_n, \xi_n) - m}{m[d(x_n, \xi_n) - 1]}$$

La Proposición 4.9 facilita el cálculo de la inversa y del determinante a partir de los calculados en el paso anterior del algoritmo. Esto aumentará la rapidez de los cálculos hechos con el ordenador.

### 6.2.3. Algoritmo acelerado

Con este proceso el algoritmo convergerá más rápido a partir de una mejor elección de  $\beta_n$  en cada paso.

1. Se elige un diseño inicial  $\xi_0$  tal que,  $\det M(\xi_0) \neq 0$ .
2. Dado un  $\xi_n$  se define un nuevo diseño:

$$\xi_{n+1} = (1 - \beta_n)\xi_n + \beta_n\xi_{x_n}$$

donde  $x_n$  y  $\beta_n$  se eligen como en el algoritmo anterior, o bien  $x_n$  es el punto donde se alcanza el mínimo de la función:

$$X \longrightarrow \Re \mid x \longrightarrow d(x, \xi_n)$$

y  $\beta_n$  es el punto donde se alcanza el máximo de:

$$\left[ \frac{-\xi_n(x_n)}{1 - \xi_n(x_n)}, 1 \right] \longrightarrow \Re \mid \beta \longrightarrow \det M[(1 - \beta)\xi_n + \beta\xi_{x_n}]$$

si  $d(x_n, \xi_n) \geq \frac{m}{1+(m-1)\xi_n(x_n)}$ , y en caso contrario  $\beta_n = -\xi_n(x_n)$ .

3. El proceso se detiene de acuerdo con la Proposición 6.2.

**Proposición 6.4** *En el proceso anterior se tiene que:*

$$\gamma_n = \frac{d(x_n, \xi_n) - m}{[(m - 1)d(x_n, \xi_n)]^{m-1}}$$

$$\det M[(1 - \beta_n)\xi_n + \beta_n\xi_{x_n}] = (1 + \gamma_n)^{-m} \det M(\xi_n)[1 + \gamma_n d(x_n, \xi_n)]$$

donde  $\beta_n(1 - \beta_n) = \gamma_n$

### 6.2.4. Algoritmo para un $X$ finito

Supongamos ahora que nos interesa encontrar el mejor diseño según el criterio D-óptimo cuyo soporte se encuentra en el conjunto  $X = \{x_1, \dots, x_k\}$ . Hay que tener en cuenta que aunque los algoritmos anteriores se dan en un conjunto compacto cualquiera,  $X$ , los cálculos del ordenador se reducen a un conjunto finito, mayor o menor, según sea la tolerancia que se asigne a los cálculos. Se propone el siguiente algoritmo:

1. Se elige un diseño inicial  $\xi_0$  tal que,  $\xi_0(x_i) > 0, i = 1, 2, \dots, k$ .
2. Dado un  $\xi_n$  se define un nuevo diseño:

$$\xi_{n+1}(x_i) = \xi_n(x_i) \frac{d(x_i, \xi_n)}{m}, i = 1, 2, \dots, k$$

3. El proceso se detiene cuando:

$$\max_i \frac{d(x_i, \xi_n)}{m} < 1 + \varepsilon$$

#### Ejemplo 6.1

Sea  $X = [0, 1]$  y el modelo:

$$E[y(x)] = \theta(x) = \alpha^t f(x) = \alpha_1 + \alpha_2 x, \sigma^2(x) = 1, x \in X$$

Es decir,  $f(x) = (1, x)^t$ . La matriz de información asociada a un diseño cualquiera,  $\xi$ , será:

$$M(\xi) = \sum_{x \in [0,1]} f(x) f^t(x) \xi(x) = \begin{pmatrix} 1 & \sum_{x \in [0,1]} x \xi(x) \\ \sum_{x \in [0,1]} x \xi(x) & \sum_{x \in [0,1]} x^2 \xi(x) \end{pmatrix}$$

Trataremos de calcular el diseño D-óptimo utilizando los algoritmos que se han enunciado anteriormente.

1. *Algoritmo de paso armónico:* Tomaremos como diseño inicial  $\xi_0 = \left\{ \begin{array}{cc} 0 & 1 \\ 1/4 & 3/4 \end{array} \right\}$ , de modo que:

$$M(\xi_0) = \begin{pmatrix} 1 & 3/4 \\ 3/4 & 3/4 \end{pmatrix}, \det M(\xi_0) = 3/16, M^{-1}(\xi_0) = \begin{pmatrix} 4 & -4 \\ -4 & 16/3 \end{pmatrix} \text{ y}$$

$$d(x, \xi_0) = \sigma^{-2}(\xi) f^t(x) M^{-1}(\xi_0) f(x) = (1, x) \begin{pmatrix} 4 & -4 \\ -4 & 16/3 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = 4 - 8x + \frac{16}{3} x^2$$

En un primer paso obtenemos el siguiente diseño:

$$\xi_1 = (1 - 1/2)\xi_0 + 1/2\xi_{x_0} = 1/2\xi_0 + 1/2\xi_{x_0}$$

donde  $x_0$  es el punto donde se alcanza el máximo de la función:

$$[0, 1] \longrightarrow \Re \mid x \longrightarrow 4 - 8x + \frac{16}{3} x^2$$

Se trata de una parábola con las ramas hacia arriba, de modo que el máximo se alcanzará en uno de los extremos del intervalo. La función toma el valor 4 para  $x = 0$  y el valor  $4/3$  para  $x = 1$ . Por tanto  $x_0 = 0$ . Utilizando la Proposición 6.2 podemos decir que el error que se cometería tomando el diseño  $\xi_0$  como D-óptimo es menor o igual que  $2 \cdot \log 4/2 = 1,39$ .

El diseño obtenido en este primer paso es entonces:

$$\xi_1 = \left\{ \begin{array}{cc} 0 & 1 \\ 5/8 & 3/8 \end{array} \right\}$$

La Proposición 4.9 nos da directamente la inversa de su matriz de información asociada:

$$\begin{aligned} M^{-1}(\xi_1) &= \frac{1}{1/2} \begin{pmatrix} 4 & -4 \\ -4 & 16/3 \end{pmatrix} - \frac{1/2 \begin{pmatrix} 4 & -4 \\ -4 & 16/3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1,0) \begin{pmatrix} 4 & -4 \\ -4 & 16/3 \end{pmatrix}}{1/2 + 4/2} = \\ &= \begin{pmatrix} 8/5 & -8/5 \\ -8/15 & 64/15 \end{pmatrix} \end{aligned}$$

Y la función:

$$[0, 1] \longrightarrow \Re \mid x \longrightarrow d(x, \xi_1) = (1, x) \begin{pmatrix} 8/5 & -8/5 \\ -8/15 & 64/15 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \frac{8}{5} - \frac{16}{5}x + \frac{64}{15}x^2$$

alcanza su máximo,  $8/3$ , en  $x_1 = 1$ . El error que se cometería tomando  $\xi_1$  como diseño D-óptimo viene acotado por la Proposición 6.2:  $2 \log(8/6) = 0,58$ .

En el siguiente paso se tomará el diseño:

$$\xi_2 = 2/3\xi_1 + 1/3\xi_{x_1} = \left\{ \begin{array}{cc} 0 & 1 \\ 5/12 & 7/12 \end{array} \right\}$$

y utilizando como antes la Proposición 4.9 calculamos la matriz inversa de la matriz de información asociada a este nuevo diseño:

$$M^{-1}(\xi_2) = \begin{pmatrix} 12/5 & -12/5 \\ -12/5 & 272/55 \end{pmatrix}$$

El máximo de la función:

$$[0, 1] \longrightarrow \Re \mid x \longrightarrow d(x, \xi_2) = \frac{12}{5} - \frac{24}{5}x + \frac{272}{55}x^2$$

se alcanza en  $x_2 = 1$  y el error cometido al tomar este diseño como D-óptimo viene acotado por  $2 \log(140/110) = 0,48$ . el siguiente diseño en este proceso sería por tanto:

$$\xi_3 = 3/4\xi_2 + 1/4\xi_{x_2} = \left\{ \begin{array}{cc} 0 & 1 \\ 5/16 & 11/16 \end{array} \right\}$$

Parece que esta sucesión de diseños tiende al diseño:

$$\xi^* = \left\{ \begin{array}{cc} 0 & 1 \\ 1/2 & 1/2 \end{array} \right\}$$

Veamos que se trata realmente del diseño D-óptimo utilizando para ello el teorema de equivalencia:

$$M(\xi^*) = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \det M(\xi^*) = 1/4, M^{-1}(\xi^*) = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix} \text{ y}$$

$$d(x, \xi^*) = 2 - 4x + 4x^2$$

cuyo máximo, 2, se alcanza en ambos extremos  $x = 0$  y  $x = 1$ . Puesto que 2 es el número de parámetros, el teorema de equivalencia nos dice que  $\xi^*$  es un diseño G-óptimo, y por tanto D-óptimo.

2. *Algoritmo ascendente:* Tomaremos el mismo diseño inicial  $\xi_0 = \left\{ \begin{matrix} 0 & 1 \\ 1/4 & 3/4 \end{matrix} \right\}$  y veremos cómo este algoritmo converge más rápidamente, aunque exige mayor número de cálculos en cada paso. Ms arriba se calcula la matriz de información inversa:

$$M^{-1}(\xi_0) = \begin{pmatrix} 4 & -4 \\ -4 & 16/3 \end{pmatrix}$$

Ahora el diseño que se obtendrá en el primer paso es:

$$\xi_1 = (1 - \beta_0)\xi_0 + \beta_0\xi_{x_0}$$

donde  $x_0$  se elige como en el algoritmo anterior, y es por tanto  $x_0 = 0$  y  $\beta_0$  es el punto donde se hace máxima la función:

$$[0, 1] \longrightarrow \Re \mid \beta \longrightarrow \det M[(1 - \beta_0)\xi_0 + \beta_0\xi_{x_0}]$$

Teniendo en cuenta que:

$$(1 - \beta_0)\xi_0 + \beta_0\xi_{x_0} = \begin{pmatrix} 0 & 1 \\ \frac{1+3\beta}{4} & \frac{1-3\beta}{4} \end{pmatrix}$$

se tiene que:

$$\det M[(1 - \beta_0)\xi_0 + \beta_0\xi_{x_0}] = \det \begin{pmatrix} 1 & \frac{3-3\beta}{4} \\ \frac{3-3\beta}{4} & \frac{3-3\beta}{4} \end{pmatrix} = \frac{3 + 6\beta - 9\beta^2}{16}$$

cuyo máximo se alcanza en  $\beta_0 = 1/3$ . Por tanto:

$$\xi_1 = \left\{ \begin{matrix} 0 & 1 \\ 1/2 & 1/2 \end{matrix} \right\}$$

que como hemos visto es el diseño D-óptimo. Esto demuestra la eficacia de este algoritmo, que en este caso consigue el diseño D-óptimo en un solo paso.

3. *Algoritmo para  $X$  finito:* Supondremos ahora que el conjunto  $X$  está formado por los once puntos que resultan de dividir el intervalo  $[0, 1]$  en 10 partes iguales, es decir:

$$X = \{0, 0,1, 0,2, \dots, 0,9, 1\}$$

Tomaremos como diseño inicial,  $\xi_0$ , el diseño uniforme concentrado en estos 11 puntos. El diseño obtenido en el primer paso vendrá dado por:

$$\xi_1(x_i) = \xi_0(x_i) \frac{d(x_i, \xi_0)}{2}, i = 0, 1, \dots, 10$$

donde se puede escribir  $x_i = i/10, i = 0, 1, \dots, 10$ . Pero:

$$M(\xi_0) = \begin{pmatrix} 1 & \sum_{i=0}^{10} \frac{i}{10} \frac{1}{11} \\ \sum_{i=0}^{10} \frac{i}{10} \frac{1}{11} & \sum_{i=0}^{10} \frac{i^2}{10^2} \frac{1}{11} \end{pmatrix} = \begin{pmatrix} 1 & \frac{55}{100} \\ \frac{55}{100} & \frac{385}{1100} \end{pmatrix}$$

$$M^{-1}(\xi_0) = \begin{pmatrix} 7/2 & -5 \\ -5 & 10 \end{pmatrix}$$

$$d(x_i, \xi_0) = (1, x_i) M^{-1}(\xi_0) \begin{pmatrix} 1 \\ x_i \end{pmatrix} = \frac{7}{2} - 10x_i + 10x_i^2 = \frac{7}{2} - i + \frac{i^2}{10}, i = 0, 1, \dots, 10$$

de modo que:

$$\xi_1(x_i) = \frac{1}{11} \frac{\frac{7}{2} - i + \frac{i^2}{10}}{2} = \frac{35 - 10i + i^2}{220}$$

que efectivamente se trata de un diseño:

$$\sum_{i=0}^{10} \xi_1(x_i) = \sum_{i=0}^{10} \frac{35 - 10i + i^2}{220} = 7/4 - 5/2 + 7/4 = 1$$

El error cometido al tomar este diseño como óptimo viene dado por:

$$\max_i \frac{d(x_i, \xi_0)}{2} - 1 = 3/4$$

De este modo se continuaría el proceso hasta lograr una buena aproximación del diseño D-óptimo o hasta apreciar cuál sería el diseño al que presumiblemente converge el proceso.

### Ejemplo 6.2

Siguiendo con el ejemplo 4.1, trataremos de buscar el diseño D-óptimo utilizando el algoritmo 8.4 para  $X$  finito. Tomaremos como diseño inicial  $\xi_0$ , es decir, el diseño uniforme concentrado en los cinco puntos de  $X$ . Como ya vimos más arriba  $\Phi_D[M(\xi_0)] = 1,02165 + 4 \log \sigma$ . En el siguiente paso obtenemos el diseño:

$$\xi_1(x_i) = \xi_0(x_i) \frac{d(x_i, \xi_0)}{m}, i = 0, 1, 2, 3, 4$$

Calculamos:

$$d(x_i, \xi_0) = \frac{5}{3} f^t(x_i) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} f(x_i) = \frac{5}{3} \|f(x_i)\|^2 = \frac{5}{3} \begin{cases} 0 & \text{si } i = 0 \\ 1 & \text{si } i = 1, 2 \\ 2 & \text{si } i = 3, 4 \end{cases}$$

Luego:

$$\xi_1 = \left\{ \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{array} \right\}$$

En la segunda iteración tendremos:

$$\xi_2(x_i) = \xi_1(x_i) \frac{d(x_i, \xi_1)}{m}, i = 0, 1, 2, 3, 4$$

Pero:

$$M(\xi_1) = \frac{1}{\sigma^2} \begin{pmatrix} 5/6 & 0 \\ 0 & 5/6 \end{pmatrix}, M^{-1}(\xi_1) = \sigma^2 \begin{pmatrix} 6/5 & 0 \\ 0 & 6/5 \end{pmatrix}$$

$$\Phi_D[M(\xi_1)] = 0,3646 + 4 \log \sigma$$

$$d(x_i, \xi_1) = f^t(x_i) \begin{pmatrix} 6/5 & 0 \\ 0 & 6/5 \end{pmatrix} f(x_i) = \frac{6}{5} \|f(x_i)\|^2 = \begin{cases} 0 & \text{si } i = 0 \\ 6/5 & \text{si } i = 1, 2 \\ 12/5 & \text{si } i = 3, 4 \end{cases}$$

De este modo obtenemos el diseño:

$$\xi_2 = \left\{ \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ 1/10 & 1/10 & 2/5 & 2/5 \end{array} \right\}$$

y su matriz de información asociada e inversa serán:

$$M(\xi_2) = \frac{1}{\sigma^2} \begin{pmatrix} 9/10 & 0 \\ 0 & 9/10 \end{pmatrix}, M^{-1}(\xi_2) = \sigma^2 \begin{pmatrix} 10/9 & 0 \\ 0 & 10/9 \end{pmatrix}$$

y el valor de la función criterio en esta matriz es:

$$\Phi_D[M(\xi_2)] = 0,2107 + 4 \log \sigma$$

$$d(x_i, \xi_2) = f^t(x_i) \begin{pmatrix} 10/9 & 0 \\ 0 & 10/9 \end{pmatrix} f(x_i) = \frac{10}{9} \|f(x_i)\|^2 = \begin{cases} 0 & \text{si } i = 0 \\ 10/9 & \text{si } i = 1, 2 \\ 20/9 & \text{si } i = 3, 4 \end{cases}$$

de modo que:

$$\xi_3 = \left\{ \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ 1/18 & 1/18 & 4/9 & 4/9 \end{array} \right\}$$

Se puede apreciar que los diseños van tendiendo ligeramente al diseño:

$$\xi^* = \left\{ \begin{array}{cc} x_3 & x_4 \\ 1/2 & 1/2 \end{array} \right\}$$

Veamos que efectivamente es el diseño D-óptimo. Para ello utilizaremos el teorema de equivalencia:

$$M(\xi^*) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, M^{-1}(\xi^*) = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$d(x_i, \xi^*) = \begin{cases} 0 & \text{si } i = 0 \\ 1 & \text{si } i = 1, 2 \\ 2 & \text{si } i = 3, 4 \end{cases}$$

de donde:

$$\max_x d(x, \xi^*) = 2 = \text{número de parámetros}$$

## Capítulo 7

# EJERCICIOS

1. Resolver los ejercicios planteados en el Capítulo 2 acerca de la inversa generalizada.
2. Sea  $X = [0, 1]$  y el modelo de regresión  $y(x) = \alpha_1 + \alpha_2 x + \varepsilon(x)$  con  $\sigma^2(x) = 1$ .
  - a) Establecer correspondencias biunívocas entre  $\mathcal{M}$  y los conjuntos:

$$\left\{ \begin{pmatrix} 1 & a \\ a & b \end{pmatrix} : 0 \leq b \leq a \leq 1, a^2 \leq b \right\} \text{ y } \left\{ M(\xi) \in \mathcal{M} : \xi = \begin{pmatrix} 0 & x_0 \\ 1-p & p \end{pmatrix} \right\}$$

- b) Demostrar que la aplicación:

$$\mathcal{M}_+ \longrightarrow \mathfrak{R} : M(\xi) \longrightarrow \text{var}_\xi \alpha_2$$

es estrictamente convexa en las matrices de información regulares.

3. Sean  $(X, \Theta_1, \sigma)$  y  $(X, \Theta_2, \sigma)$  dos modelos de regresión. Supongamos que  $\Theta_1 \subset \Theta_2$  y que  $\xi$  es un diseño en ambos modelos. Sea  $g$  un funcional lineal en  $\Theta_2$  y sea  $g|_{\Theta_1}$  la restricción de  $g$  al conjunto  $\Theta_1$ . Probar que:

$$\text{var}_\xi g|_{\Theta_1} \leq \text{var}_\xi g$$

Indicación: Utilícese la expresión de la varianza del BLUE.

4. Se tiene un polinomio de grado  $m$  con coeficientes desconocidos. Se puede observar su valor, independientemente, en  $N$  puntos  $x_1, \dots, x_N \in [0, 1]$ . Se supone que  $\sigma^2(x) = \sigma^2$  es constante. Encontrar una expresión reducida del determinante de la matriz de información del diseño exacto  $x_1, \dots, x_N$  en el caso  $N = m + 1$ .
5. Sea  $(X, \Theta, \sigma)$  un modelo de regresión lineal con observaciones incorreladas. Sean  $x_1, \dots, x_k \in X$  y  $\alpha_1, \dots, \alpha_k \in [0, 1]$  tales que  $\sum_{i=1}^k \alpha_i < 1$ . Sea  $\Phi$  una función criterio y sea  $\xi^*$  un diseño  $\Phi$ -óptimo. Denotaremos:

$$\Xi_S = \{ \xi \in \Xi : \xi(x_1) \geq \alpha_1, \dots, \xi(x_k) \geq \alpha_k \}$$

$$\mathcal{M}_S = \{ M(\xi) \in \mathcal{M} : \xi \in \Xi_S \}$$

- a) Demostrar que  $\mathcal{M}$  y  $\mathcal{M}_S$  son homeomorfos. Deducir que si  $\mathcal{M}$  es compacto entonces también lo será  $\mathcal{M}_S$ .

- b) Encontrar una relación entre las matrices de información  $M(\xi) \in \mathcal{M}$  y su imagen,  $M(\xi_0) \in \mathcal{M}_S$ , por el homeomorfismo natural dado en el primer apartado.
- c) Deducir del segundo apartado que el diseño imagen  $\xi_0^*$  del diseño  $\xi^*$  verifica que:

$$\Phi[M(\xi_0^*)] = \max_{\xi \in \Xi} \Phi[M(\xi^*)]$$

6. Sea  $X = [0, 1]$  y el modelo de regresión  $y(x) = \alpha e^{-\beta x^2} + \varepsilon(x)$  con  $\sigma^2(x) = 1$ . Por la experiencia anterior se obtiene un primera estimación de los parámetros  $\alpha_0 = 1$  y  $\beta_0 = 1/2$ .

- a) Linealizar el modelo.
- b) Encontrar la matriz de información del diseño:

$$\xi_0 = \left\{ \begin{array}{cc} 0 & 1 \\ 1/2 & 1/2 \end{array} \right\}$$

- c) Demostrar que no es un diseño D-óptimo utilizando el teorema de equivalencia y acotar el error cometido al tomar este diseño como D-óptimo.

# Bibliografía

- [1] ABDELBASIT K. M. y PLACKETT R. L. (1983). Experimental design for binary data. *Journal of the american statistical association*, Vol. 78, No. 381, pp 90-98.
- [2] ABRAMOVITZ M. y STEGUN I. A. (1972). *Handbook of mathematical functions*. Dover Pub. Inc., New York, pp 773-801.
- [3] AGARWAL S.C. and DAS M.N. (1990). Incomplete Blocks Designs for Partial Diallel Cross. *Sankhya*, Vol. 52, Series B, Pt. 1, pp 75-81.
- [4] AIGNER D. J. y BALESTRA P. (1988). Optimal experimental design for error components models. *Econometrica*, Vol. 56, No. 4, pp 955-971.
- [5] ALBERT A. (1972). *Regression and the Moore-Penrose pseudoinverse*. Academic Press. Ser. Mathematics in science and engineering, Vol 94.
- [6] ALI M. A. y SINGH N. (1981). Optimal use of prior information in designing experiments. *Journal of information and optimization sciences*, Vol. 2 No. 1 pp 64-72.
- [7] ASH A. y HEDAYAT A. (1978). An introduction to design optimality with an overview of the literature. *Commun. statist. -theor. meth., Ser. A*, Vol. 7, No. 14, pp 1295-1325.
- [8] ATKINSON A. C. (1978). Posterior probabilities for choosing a regression model. *Biometrika*, Vol. 65, pp 39-48.
- [9] ATKINSON A. C. (1982). Developments in the design of experiments. *International Statistical review*, Vol. 50, pp 161-177.
- [10] ATKINSON A.C. y DONEV A.N. (1992). *Optimum Experimental Designs*. Oxford Science Publications, New York.
- [11] ATKINSON A.C. y FEDOROV V.V. (1975). The design of experiments for discriminating between two rival models. *Biometrika*, Vol. 62, No. 1, pp 57-70.
- [12] ATKINSON A. C. y FEDOROV V. V. (1975). Optimal design: Experiments for discriminating between several models. *Biometrika*, Vol. 62, No. 2, pp 289-303.
- [13] ATKINSON A. C. y FIENBERG S. E. (Eds.) (1985). *An introduction to the optimum design of experiments. A Celebration of statistics*. Springer-Verlag. pp 465-473.
- [14] ATKINSON A. C. y HUNTER W. G. (1968). The design of experiments for parameter estimation. *Technometrics*, Vol. 10, No. 2, pp 271-289.

- [15] ATWOOD C. L. (1969). Optimal and efficient designs of experiments. *The annals of mathematical statistics*, Vol. 40, No. 5, pp 1570-1602.
- [16] ATWOOD C. L. (1973). Sequences Converging to D-Optimal Designs of Experiments. *The Annals of Statistics*, Vol. 1, No. 2, pp 342-352.
- [17] ATWOOD C. L. (1976). Convergent design sequences, for sufficiently regular optimality criteria. *The annals of statistics*, Vol. 4, No. 6, pp 1124-1138.
- [18] ATWOOD C. L. (1980). Convergent design sequences, for sufficiently regular optimality criteria, II: singular case. *The annals of statistics*, Vol. 8, No. 4, pp 894-912.
- [19] BAGCHI S. (1986). A series of nearly D-optimal third order rotatable designs. *The indian journal of statistics*, Vol. 48, series B, No. 2, pp 186-198.
- [20] BAGCHI S., MUKHOPADHYAY A.C. and SINHA B.K. (1990). A Search for Optimal Nested Row-Column Designs. *Sankhya*, Vol. 52, Series B, Pt. 1, pp 93-104.
- [21] BANDEMER H. (1980). Problems in foundation and use of optimal experimental design in regression models. *Math. Operationsforsch. Statistics., Ser. Statistics*, Vol.11, No. 1, pp 89-113.
- [22] BANDEMER H. y NAGEL W. (1981). Parameter estimation in linear regression models with weak and fuzzy prior knowledge. *Math. Operationsforsch. Statist., Ser. Statistics*, Vol. 12, No. 3, pp 297-305.
- [23] BANDEMER H., NATHER W. y PILZ J. (1987). Once more: Optimal experimental design for regression models (with discussion). *Statistics*, Vol.18, No. 2, pp 171-217.
- [24] BARDESLEY W. G., MACGINLAY P. B. y ROIG M. G. (1987). Conditions when statistics test for model discrimination have high power. Some examples from pharmacokinetics, ligand binding, transient and steady-state enzyme kinetics. *Biophysical Chemistry*, Vol. 26, pp 1-8.
- [25] BARDESLEY W. G., MACGINLAY P. B. y ROIG M. G. (1989). Optimal design for model discrimination using the F-test with non-linear biochemics models. Criteria for choosing the number and spacing of experimental points. *J. theor. biol.*, Vol. 139, pp 85-102.
- [26] BARDESLEY W. G., MACGINLAY P. B. y WRIGHT A. J. (1986). The F-test for model discrimination with exponential functions. *Biometrika*, Vol. 37, No.2, pp 501-508.
- [27] BARDESLEY W. G. y WAIGHT R. D. (1978). Factorability of the hessian of the binding polynomial. The central issue concerning statistical ratios between binding constants, hill plot slope and positive and negative co-operativity. *J. theor Biol.*, Vol. 72, pp 321-372.
- [28] BARDESLEY W. G., WOOLFSON R. y WOOD R. M. W. (1980). Some mathematical results concerning Hessians of binding polynomials and Co-operativity coefficients. *J. Theor. Biol.*, Vol. 85, No. 1, pp 45-51.
- [29] BATES D. (1983). The derivative of  $|X^T X|$  and its uses. *Technometrics*, Vol. 25, No. 4, pp 373-376.
- [30] BAYARD D. S., HADAEGH F. Y. y MELDRUM D. R. (1988). Optimal experimental design for identification of large space structures. *Automatica*, Vol. 24, No. 3, pp 357-364.

- [31] BECHER H. (1988). On optimal experimental design under spatial correlation structures for square and nonsquare plot designs. *Commun. statist. -simula.*, Vol 17, No. 3, pp 771-780.
- [32] BELLHOUSE D.R. y HERZBERG A.M. (1984). Equally spaced desing points in polynomial regression: a comparison of systematic sampling methods with the optimal desing of experiments. *The canadian journal of statistics*, Vol.12, No.2, pp 77-90.
- [33] BICKEL P. J. (1973). On some analogues to linear combinations of order statistics in the linear model. *The annals os statistics*. Vol. 1, No. 4,pp 597-616.
- [34] BIEMBER P. P. y STOKES S. L. (1985). Optimal design of interviewer variance experiments in complex surveys. *Journal of the american statistical association*, Vol. 80, No. 389, pp 158-166.
- [35] BOHNING D. (1981a). On the construction of optimal experimental designs: a penalty approach. *Math. Operationsforsch. statist., ser. statistics*, Vol. 12, No. 4, pp 487-495.
- [36] BOHNING D. (1981b). On an assumption of Bandemer and Ketzel in optimal experimental design theory. *Math. operationsforsch. statist., Ser. statistics*, Vol. 12, No. 4, pp 497-502.
- [37] BOHNING D. (1984). Use of reparameterization in nonlinear optimization with applications to statistics and optimal design. *Computational statistics quarterly*. Vol. 1, No. 1, pp 29-43.
- [38] BOX G. E. P. (1979). Some problems of statistics and everyday life. *Journal of the american statistical association*, Vol. 74, pp 1-4.
- [39] BOX G. E. P. y DRAPER N. R. (1959). A basis for the selection of a response surface design. *Journal of the american statistical association*, Vol. 54, pp 622-654.
- [40] BOX G. E. P. y HUNTER W. G. (1965). Sequential design of experiments for nonlinear models. *IBM Scientific computing symposium in statist.*, pp113-137.
- [41] BOX G. E. P. y LUCAS H. L. (1959). Desing of experiments in non-linear situations. *Biometrika*, Vol. 46, pp 77-90.
- [42] BOX M. J. (1968a). The use of designed experiments in nonlinear model building. *The future of statist*, pp 241-268.
- [43] BOX M. J. (1968b). The occurrence of replications in optimal designs of experiments to estimate parameters in nonlinear models. *Journal royal statistical society, Ser. B*, Vol. 30, pp 290-302.
- [44] BOX M. J. (1969). The design of experiments and nonlinear model-building under a variety of error estructures. Ph. D. Thesis, University of London, London, Englan.
- [45] BOX M. J. (1970). Some experinces with a nonlinear experimental design criterion. *Technometrics*, Vol. 12, pp 569-589.
- [46] BOX M. J. y DRAPER N. R. (1971). Factorial designs, the  $|X'X|$  criterion, and some related matters. *Technometrics*, Vol. 13, No. 4, pp731-742.
- [47] BRILLINGUER D. R. (1975). *Time series*. Holt, Rinehart and Winston, Inc., pp 70-87.

- [48] BROOKS R. J. (1972). A decision theory approach to optimal regression designs. *Biometrika*, Vol. 59, No. 3, p 563.
- [49] BROOKS R. J. (1976). Optimal regression designs for prediction when prior knowledge is available. *Metrika*, Band 23, pp 221-230.
- [50] BROWN L. D., OLKIN I., SACKS J. y WYNN H. P. (Eds) (1985). *Jack Carl Kiefer Collected Papers III, Design of Experiments*. Springer Verlag. New York.
- [51] CALVETE H. I. (1983). *Diseño óptimo de experimentos: Diseños marginalmente restringidos*. Tesis doctoral, Facultad de Ciencias, Univ. de Zaragoza.
- [52] CALVIN J. A. (1986). A new class of the covariance balanced designs. *Journal of statistical planning and inference*, Vol. 14, No. 2, pp 251-254.
- [53] CHAN L-Y. (1988). Optimal design for a linear log contrast model for experiments with mixtures. *Journal of statistical planning and inference*, Vol. 20, pp 105-113.
- [54] CHENG CH.-S. (1978a). Optimality of certain asymmetrical experimental designs. *The annals of statistics*, Vol. 6, No. 6, pp1239-1261.
- [55] CHENG CH.-S. (1978b). Optimal design for the elimination of multi-way heterogeneity. *The annals of statistics*, Vol 6, No. 6, pp 1262-1272.
- [56] CHERNOFF H. (1953). Locally optimal designs for estimating parameters. *The annals of mathematical statistics*, Vol. 24, pp 586-602.
- [57] CLARKE G. M. (1980). *Statistics and experimental design*. Ed. Edward Arnold Ltd.
- [58] COSTA LOPEZ J., MATA ALVAREZ J. y DOMINGO CAMPOS X. (1981). Diseños experimentales: Introducción y diseños axiales. *Quaderns d'enginyeria*, Vol. 3, No. 1, pp 279-285.
- [59] COVEY-CRUMP P. A. K. y SILVEY S. D. (1970). optimal regression designs with previous observations. *Biometrika*, Vol. 57, No. 3, pp 551-566.
- [60] CURRIE D. J. (1982). Estimating Michaelis-Menten Parameters: bias, variance and experimental design. *Biometrics*, Vol. 38, pp 907-919.
- [61] DE A.K. and ROY B.K. (1990). Computer Construction of Some Group Divisible Designs. *Sankhya*, Vol. 52, Series B, Pt. 1, pp 82-92.
- [62] DE LA GARZA A. (1954). Spacing of information in polynomial regression. *The annals of mathematical statistics*, Vol. 25, pp 123-130.
- [63] DETTE H. (1993). On the Mixture of the D- and D1-optimality criterion in polinomial regression. *Journal of Statistical Planing and Infeerence*, Vol. 35, pp 233-249.
- [64] DETTE H. y STUDDEN W. J. (1993). *Minimax Designs in Linear Regression Models*. Preprint.
- [65] DODGE, FEDOROV y WYNN (Eds.) (1988). *Optimal design and analysis of experiments*. First international conference workshop on optimal desig and analysis experiments. Elsevier science publishers B. V.

- [66] DRAPER N. R. y HUNTER W. G. (1967). The use of prior informations in the design of experiments for parameter estimation in non-linear situations: multiresponse case. *Biometrika*, Vol. 54, pp 662-665.
- [67] ECCLESTON J. A. y JONES B. (1980). Exchange and interchange procedures to search for optimal row-and-column designs. *Journal royal statistical society, Ser. B*, Vol. 42 No. 3 pp 372-376.
- [68] EHRENFELD S. (1955). On the efficiency of experimental designs. *The annals of mathematical statistics*, Vol 26, pp 247-255.
- [69] EHRENFELD S. (1956). Complete Class Theorems in Experimental Design. *Proc. Third Berkeley Symp. Math. Statist. Probab.*, Vol. 1, pp 57-67. Univ California Press, Berkeley.
- [70] ELFVING G. (1952). Optimum allocation in linear regression theory. *The annals of mathematical statistics*, Vol. 23, pp 255-262.
- [71] ELFVING G. (1959). *Designs of Linear Experiments. Prob. and Stats. The Harold Cramér Volume*. Ed. Ulf Grenander Wiley, pp 58-74.
- [72] ENDRENYI L. y CHAN F.-Y. (1981). Optimal design of experiments for the estimation of precise hiperbolic kinetic and binding parameters. *J. theor. Biol.*, Vol. 90, pp 241-236.
- [73] FEDERER W. T. (1980). Some recent results in experiment design whit a bibliography - I. *International statistical review*, Vol. 48, pp 357-368.
- [74] FEDOROV V.V. (1972). *Theory of optimal experiments*. Department of statistics Purdue University.
- [75] FEDOROV V. V. (1975). Optimal Experimental Design for Discriminating two Rival Regression Models. J. N. Srivastava, ed., *A Survey of Statistical Design and Linear Models*. North-Holand Publishing Company, pp 155-164.
- [76] FEDOROV V. V. (1980). Convex design theory. *Math. Operationsforsch. Statist., Ser. statistics.*, Vol. 11, No. 3, pp 403-413.
- [77] FEDOROV V. V. (1988). Properties of optimal designs of regression experiments in singular cases. *Journal of soviet mathematics*, Vol. 40, No. 2, pp 245-246.
- [78] FEDOROV V. V. y KHABAROV V. (1986). Duality of optimal designs for model discrimination and parameter estimation. *Biometrika*, Vol. 73, No. 1, pp 183-190.
- [79] FELLMAN J. (1974). On the allocation of linear observations. *Commentationes Physico-Mathematicae*, Helsinki, Vol. 44, pp 27-78.
- [80] FLEMING W. H. (1965). *Functions of several variables*. Addison-Wesley Pub. Co., Inc., Reading, Massachusetts, pp 22-29 y 52-61.
- [81] FORD I. (1976). Extract from Thesis. Ph.D. Thesis. University of Glasgow.
- [82] FORD I., TITTERINGTON D. M. and KITSOS CH. P. (1989). Recent Advances in Nonlinear Experimental Design. *Technometrics*, Vol. 31, No. 1, pp 49-60.

- [83] FORD I., TORSNEY B. y WU C.F.J. (1992). The Use of a Canonical Form in the Construction of Locally Optimal Designs for Non-Linear Problems. *J. R. Statist. Soc., Ser. B*, Vol. 54, No. 2, pp 569-583.
- [84] FRISCHMUTH K. (1989). On locally D-optimal experimental design. *Rostock.math. kolloq.*, Vol. 36, pp 51-64.
- [85] GALIL Z. y KIEFER J. (1980a). D-optimum weighing designs. *The annals of statistics*, Vol. 8, No. 6, pp 1293-1306.
- [86] GALIL Z. y KIEFER J. (1980b). Time-and space- saving computer methods, related to Mitchell's DETMAX, for finding d-optimum designs. *Technometrics*, Vol. 22, No. 3, pp 301-313.
- [87] GALIL Z. y KIEFER J. (1982). Construction methods for D-optimum weighing designs when  $n=3(\text{mod}4)$ . *The Annals of Statistics*, Vol. 10, No. 2, pp 502-510.
- [88] GALIL Z. y KIEFER J. (1983). Comparison of designs equivalent under one or two criteria. *Journal of statistical planning and inference*, Vol 8, pp 103-116.
- [89] GEVERS M. y LJUNG (1986). Optimal experimental design with respect to the intended model application. *Automatica*, Vol. 22, No. 25, pp543-554.
- [90] GHOSH S. (1986). On a new graphical method of determining the connectedness in three dimensional designs. *The indian journal of statistics*, Vol. 48, Ser. B, Pt. 2, pp 207-215.
- [91] GILL E. P., MURRAY W. y WRIGHT M. H. (1981). *Practical Optimization*. Academic Press.
- [92] GLADITZ J. y PLIZ J. (1982). Construction of optimal designs in random coefficient regression models. *Math. Operationsforsch. Statistics*, Vol. 13, No. 3, pp 371-365.
- [93] GOLŠTEIN E. G. (1972). *Theory of convex programming*. American Mathematical Society providence. Rhode Island.
- [94] GRAFFKE N. (1985). Directional derivatives of optimality criteria at singular matrices in convex design theory. *Statistics*, Vol. 16, No. 3, pp 373-388.
- [95] GRAYBILL F. A. (1983). *Matrices with applications in statistics*. Wadsworth.
- [96] GUEST P. G. (1958). The spacing of observations in polynomial regression. *The annals of mathematical statistics*, Vol 29, pp 294-299.
- [97] GUPTA R. y RICHARDS D. (1985). Testing optimality of experimental designs for a regression model with random variables. *Journal of statistical planning and inference*, Vol. 11, pp 75-80.
- [98] HAHN G. J. (1984). Experimental design in the complex world. *Technometrics*, Vol. 26, No. 1, pp 19-31.
- [99] HARDY G., LITTLEWOOD J. E. y POLYA G. (1989). *Inequalities*. Cambridge university press. New York, Port Chester, Melbourne y Sydney.

- [100] HARVILLE D. A. (1975). Computing Optimum Design for Covariance Models. J. N. Srivastava, ed., A Survey of Statistical Design and Linear Models. North-Holland Publishing Company, pp 209-228.
- [101] HEALY M. J. R. (1986). Matrices for statistics. Oxford science publications.
- [102] HEDAYAT A. (1981). Study of optimality criteria in design of experiments. Statistics and related topics. North Holland Publishing Company, pp 39-56.
- [103] HERZBERG A. M. (1982). The robust design of experiments: A review. *Serdica Bulgaricae mathematicae publicationes*, Vol. 8, pp 223-228.
- [104] HERZBERG A. M. y ANDREWS D. F. (1973). Some considerations in the optimal design of experiments in non-optimal situations. *Journal royal statistical society, Ser. B* 8, No. 3, pp 284-289.
- [105] HERZBERG A. M. y TSUKANOV A. V. (1985). The design of experiments for linear model selection with the jackknife criterion. *Utilitas mathematica*, Vol. 28, pp 243-253.
- [106] HILL D. H. (1980). D-optimal designs for partially nonlinear regression models. *Technometrics*, Vol. 22, No. 2, pp 275-276.
- [107] HOEL P. G. (1958). Efficiency problems in polynomial estimation. *The annals of mathematical statistics*, Vol. 29, pp 1134-1145.
- [108] HOEL P. G. y JENNRICH R. I. (1979). Optimal design for dose response experiments in cancer research. *Biometrika*, Vol. 66, No. 2, pp 307-316.
- [109] HOEL P. G. y LEVINE A. (1964). Optimal spacing and weighting in polynomial prediction. *The annals of mathematical statistics*, Vol. 35, pp 1553-1560.
- [110] HOLMES R. B. (1975). *Geometric functional analysis and its applications*. Springer, New York.
- [111] HUBER P. J. (1973). The 1972 wald memorial lectures. *The annals of statistics*, Vol. 1, No. 5, pp 799-821.
- [112] INTRILIGATOR M. D. (1971). *Mathematical optimization and economic theory*. Prentice-Hall series in mathematical economics.
- [113] JACROUX M. (1983). On the MV-Optimality of Chemical Balance Weighing Designs. *Calcutta Statist. Assoc. Bull.* Vol 32, pp 143-151.
- [114] JACROUX M. (1987). On the Determination and Construction of MV-optimal Block Designs for Comparing Test Treatments with a Standard Treatment. *Journal of Statistical Planning and Inference*, Vol. 15, pp 205-225.
- [115] JACROUX M. (1988). Some Further Results on the MV-Optimality of Block Designs for Comparing Test Treatments to a Standard Treatment. *J. Statist. Plann. Inference*, Vol. 20, pp 201-214.
- [116] JOHN J. A. y QUENOVILE M. H. (1977). *Experiments: design and analysis*. Ed. Charles Griffin and Company Ltd.

- [117] JOHN R. C. ST. y DRAPER N. R. (1975). D-optimality for regression designs: A review. *Technometrics*, Vol. 17, No.1, pp 15-23.
- [118] JURECKOVA J. (1971). Nonparametric estimate of regression coefficients. *The annals of mathematical statistics*, Vol.42, pp 1328-1338.
- [119] KAISHEV V. K. (1989). Optimal experimental designs for the B-spline regression. *Computational statistics and data analysis*, Vol. 8, pp 39-47.
- [120] KARLIN S. y STUDDEN W. (1966). Optimal experimental designs. *The annals of mathematical statistics*, Vol. 37, pp 783-810.
- [121] KAZIMIERIERCZYK P. (1989). Optimal experimental design beam under random loading. *European journal of mechanics a/solids*, Vol. 8, No.3, pp 161-184.
- [122] KIEFER J. (1959). Optimum experimental designs. *Journal royal statistical society, Ser. B*, Vol. 21, pp 272-319.
- [123] KIEFER J. (1961). Optimum designs in regression problems, II. *The annals of mathematical statistics*, Vol 32, pp 298-325.
- [124] KIEFER J. (1974). General equivalence theory for optimum designs (Approximate theory). *The annals of statistics*, Vol. 2, No.5, pp 848-879.
- [125] KIEFER J. (1975a). Optimal design: Variation in structure and performance under change of criterion. *Biometrika*, Vol. 62, No. 2, p 277-288.
- [126] KIEFER J. (1975b). Construction and Optimality of Generalized Youden Designs. J. N. Srivastava, ed., *A Survey of Statistical Design and Linear Models*. North-Holland Publishing Company, pp 333-353.
- [127] KIEFER J. y STUDDEN W. J. (1976). Optimal design for large degree polynomial regression. *The annals of statistics*, Vol. 4, No. 6, pp 1113-1123.
- [128] KIEFER J. y WOLFOWITZ J. (1959). Optimum design in regression problems. *The annals of mathematical statistics*, Vol. 30, pp 271-294.
- [129] KIEFER J. y WOLFOWITZ J. (1960). The equivalence of two extremum problems. *Canad. J. Math.*, Vol 12, pp 363-366.
- [130] KITSOS C. P. (1992). Quasi-Sequential Procedures for the Calibration Problem. *Proceedings of the 10th Symposium on Computational Statistics*, pp 227-231. Physica-Verlag. Neuchatel, Switzerland.
- [131] KITSOS C. P., TITTERINGTON D. M. and TORSNEY B. (1988). An Optimal Design Problem in Rhythmometry. *Biometrics*, Vol. 44, pp 657-671.
- [132] KOCH G. G. y KUMAR P. (1968). Some aspects of the statistical analysis of the mixed model. *Biometrics*, March, pp 27-48.
- [133] KRAFFT O. (1983). A matrix optimization problem. Elsevier Science publishing Co., Inc., pp 137-142.

- [134] KRESWSKI D., BICKIS M., KOVAR J. y ARNOLD D. L. (1986). Optimal experimental design for low dose extrapolation I. The case of zero background. *Utilitas mathematica*, Vol. 29, pp 245-262.
- [135] LANDAW E. M. (1985). Optimal design for individual parameter estimation in Pharmacokinetics. M Rowland et al. Raven Press, pp 187-199.
- [136] LAYCKOCK P. J. and SILVEY S. D. (1968). Optimal designs in regression problems with a general convex loss function. *Biometrika*, Vol. 55, No. 1, pp 53-66.
- [137] LAYCOCK P. J. y ROWLEY P. J. (1991). An Algorithm for Generation of all Regular Fractions and Blocks. Technical Report No. 204, Department of Mathematics, UMIST, Manchester.
- [138] LAYCOCK P. J. (1972). Convex Lost Applied to Design in Regression Problems. *J. R. Statist. Soc., Series B*, Vol. 34, pp 148-170.
- [139] LAYCOCK P. J. y ROWLEY P. J. (1990). On Generating and Classifying all Regular Fractions and Blocks for  $q^{n-p}$  Designs. Technical Report No. 201, Department of Mathematics, UMIST, Manchester.
- [140] LAYCOCK P. J. y SEIDEN E. (1980). On a problem of Repeated Measurement Design with Treatment Additivity. *The Annals of Statistics*, Vol. 8, No. 6, pp 1284-1292.
- [141] LEAMER E. E. (1982). Sets of posterior means with bounded variance priors. *Econometria*, Vol. 50, No. 3, pp 725-736.
- [142] LIM Y. B., STUDENT W. J. y WYNN H. P. (1988). A note on aproximate D-optimal designs for Gx2M. Papers from the fourth Purdue Symposium on Statistical Decision and related Topics, held at Purdue University, June 15-20, pp 351-361, 1986.
- [143] LOPES TROYA J. (1982). Optimal designs for covariates models. *Journal of statistical planning and inference*, Vol. 6, pp 373-419.
- [144] MANDAL N. K. (1982). D-optimal designs for stimating the optimum point in a multifactor experiment. *Calcutta statist. assoc. bull.*, Vol 31, No. 123-124, pp 105-130.
- [145] MEENA N. S. (1980). Optimum spring balance weighing designs for estimating the total weight. *Commun. statist.-theor. meth.*, Ser. A, Vol. 9, No. 11, pp 1185-1190.
- [146] MEHRA K. L. y SMITH G. E. J. (1970). On nonparametric estimation and testing for interactions in factorial experiments. *Journal of the american statistical association*, Vol. 65, No 331, pp 1283-1296.
- [147] MEYER R. K. y NACHTSHEIM CH. J. (1988). Constructing exact D-optimal experimental designs by simulated annealing. *American journal of mathematical and management sciences*, Vol. 8, Nos. 3 y 4, pp 329-359.
- [148] MINKIN S. (1987). Optimal designs for binary data. *Journal of the american statistical association*, Vol.82, No 4000, pp 1098-1103.
- [149] MITCHELL T. J. (1974). An algorithm for the construction of "D-optimal" experimental designs. *Technometrics*, Vol. 16, No. 2, pp 203-210.

- [150] MOLINA SORIANO R. y FERNADEZ VIVAS C. (1987). Una caracterización de diseños de experimentos óptimos por subgradientes. *Trabajos de estadística*, Vol. 2, No. 1, pp 3-13.
- [151] MUKERJEE R. y HUDA S. (1985). D-Optimal statistical designs with restricted string property. *Commun. Statist.-Theor. Meth.*, Vol. 14 (3), pp 669-677.
- [152] MULLER-FUNK U., PUKELSHEIM F. y WITTING H. (1985). On the duality between locally optimal tests and optimal experimental designs. *Linear algebra and its applications*, Vol. 67, pp 19-34.
- [153] MULLER W. G. (1991). Optimal Design for Moving Local Regressions. *Forschungsbericht/Research Memorandum No. 281*. Institute for Advanced Studies, Vienna.
- [154] MULLER W. G. (1992). Optimal Design for Moving Local Regressions. *Institut für Statistik*, No. 29.
- [155] NAGEL S. (1978). Goodness of exact experimental designs with respect to a certain family of criteria. *Math. Operationsforsch. Statist., Ser. statistics*, Vol. 9, No. 4, pp 519-523.
- [156] NALIMOV V. V., GOLIKOVA T. I. y MIKESHINA N. G. (1970). On practical use of the concept of D-optimality. *Technometrics*, Vol. 12, No. 4.
- [157] NATHER W. y PLIZ J. (1980). Estimation and experimental design in a linear regression model using prior information. *Zastosowania matematyki applicationes mathematicae XVI*, Vol. 4, pp 565-577.
- [158] NATHER W. y REINSCH V. (1981). Ds-optimality and Whittle's equivalence theorem. *Math. Operationsforsch. Statist., Ser. statistics*, Vol. 12, No. 3, pp 307-316.
- [159] NELSON P. R. (1988). Testing for interactions using the analysis of means. *Technometrics*, Vol 30, No. 1, pp 53-61.
- [160] NG T. S. y QURESHI Z. H. (1981). Optimal experimental design for autoregressive model with output power constraints. *IEEE Transactions on automatic control*, Vol. AC26, No. 3, pp 739-742.
- [161] NIGAM A. K., GUPTA S. C. y GUPTA S. (1983). A new algorithm for extreme vertices designs for linear mixture models. *Technometrics*, Vol. 25, No. 4, pp 367-371.
- [162] O'HAGAN A. (1978). Curve fitting and optimal design for prediction. *Journal royal statistical society, Ser. B*, No. 40, pp 1-42.
- [163] OWEN R.J. (1970). The optimum design of a two-factor experiment using prior information. *The annals of mathematical statistics*, Vol. 41, No. 6, pp 1917-1934.
- [164] PAPAKIRIAZIS P. A. (1978). Optimal experimental design in econometrics. The time series problem. *Journal of econometrics*, Vol. 7, pp 351-372.
- [165] PAZMAN A. (1974). A Convergence Theorem in the Theory of D-Optimum Experimental Designs. *The Annals of Statistics*, Vol. 2, No. 1, pp 216-218.
- [166] PAZMAN A. (1980a). Singular experimental designs (Standard and Hilbert-space approaches). *Math. operationsforsch. statist., Ser. statistics*, Vol. 11, No. 1, pp 137-149.

- [167] PAZMAN A. (1980b). Some features of the optimal designs theory -A survey. *Math. Operationsforsch. Statist., Ser. statistics*, Vol. 11, No. 3, pp 415-446.
- [168] PAZMAN A. (1986). *Foundations of optimum experimental design*. D. Reidel publishing company. Dordrecht.
- [169] PAZMAN A. (1990). Small-Sample Distributional Properties of Nonlinear Regression Estimators (a Geometric Approach). *Statistics*, Vol. 21, No. 3, pp 323-367.
- [170] PAZMAN A. and PRONZATO L. (1992). Second-Order Aproximation of the Entropy in Nonlinear Least-Squares Estimation. *Rapport interne L2S/92*.
- [171] PUKELSHEIM F. (1981). On  $c$ -Optimal Design measures. *Math. Operationsforsch. Statist., Ser. Statistics*, Vol. 12, No.1, pp13-20.
- [172] PUKELSHEIM F. (1980). On linear regression designs wich maximize information. *Journal of statistical planning and inference*, Vol. 4, pp 339-346.
- [173] PUKELSHEIM F. (1993). *Optimal Design of Experiments*. John Wiley and Sons. New York.
- [174] PUKELSHEIM F. y TITTERINGTON D. M. (1983). General diferential and lagrangian theory for optimal esperimental design. *The annals of statistics*, Vol. 11, No. 4, pp 1060-1068.
- [175] PUKELSHEIM F. y TORSNEY B. (1991). Optimal Weights for Experimental Designs on Linearly Independent Support Points. *The Annals of Statistics*, Vol. 19, No. 3, pp 1614-1625.
- [176] RAO C. R. (1962). A note on a generalized inverse of a matrix with aplications to problems in matematical statistics. *Journal royal statistical society, Ser. B*, Vol. 24, pp 152-158.
- [177] RASCH D. y HERRENDORFER G. (1986). *Experimental designs: Sample Size determinations and block designs*. Ed. D. Reidel Publishing Company., Dordrecht-Boston, Mass.
- [178] ROWLEY P. J. y LAYCOCK P. J. (1991). *Minimum Aberration Designs*. Tecnical Report No. 201, Departmentof Mathematics, UMIST, Manchester.
- [179] SCHMIDT W. H. (1977). Asymptotics in multivariate linear models with optimal experimental design. *Math. operationsforsch. statist., Ser. Statistics*, Vol. 8, No. 4, pp 447-452.
- [180] SCHOENBERG L. J. (1959). On the maxima of certain hankel determinants and the zeros of the classical orthogonal polynomials. *Nederl. Akad. Netensch. Proc., Ser. A 62*, Vol. 21, pp 282-290.
- [181] SEN M. y MUKERJEE R. (1987). Optimal repeatead measurements designs under interaction. *Journal of statistical planing and inference*, Vol. 17, pp 81-91.
- [182] SHISHA O. (1967). *Inequalities*. Academic Press.
- [183] SIBSON R. (1973). DA-optimality and duality. *Colloq. math. soc. Janos Bolyai*, Vol. 9 II, pp 677-692, North Holland.
- [184] SILVEY S. D. (1974). Some aspects of the theory of optimal linear regression design with a general concave criterion function. *Techn. rep., No. 75, Ser. 2*. Department of statistics, Princenton university.

- [185] SILVEY S. D. (1978). Optimal design measures with singular information matrices. *Biometrika*, Vol. 65, No. 3, pp 553-9.
- [186] SILVEY S.D. (1980). Optimal design. Chapman and Hall.
- [187] SILVEY S. D. y TITTERINGTON D. M. (1973). A geometric approach to optimal design theory. *Biometrika*, Vol. 60, No. 1, pp 21-32.
- [188] SINGH N. y PEIRIS M. S. (1987). Optimal experimental designs for linear time series models with stochastic coefficients. *J. indian soc. stat. ops. res.*, Vol. 8, No. 1-4, pp 1-9.
- [189] SMITH K. (1918). On the standard deviations of adjusted and interpolates values of an observed polynomial functions and its constanst and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, Vol. 12, pp 1-85.
- [190] SNEE R. D. y MARQUARDT W. (1974). Extreme vertices designs for linear mixture models. *Technometrics*, Vol. 16, No. 3, pp 399-408.
- [191] SOARES J. F. y WU C. F. J. (1985). Optimality of random allocation design for the control of accidental bias in sequential experiments. *Journal of statistical planning and inference*, Vol. 11, pp 81-87.
- [192] SPRUILL M.C. (1990). Good Designs for Testing the Degree of a Polynomial Mean. *Sankhya*, Vol. 52, Series B, Pt. 1, pp 67-74.
- [193] SRIVASTAVA J. N. y WIJETUNGA A. M. (1983). D-óptimal cyclic two-dimensional block designs. *Trabajos de estadística y de investigación operativa*, Vol. 34, No. 3, pp 95-108.
- [194] STEINBERG D. M. y HUNTER W. G. (1984). Experimental design: Review and coment. *Technometrics*, Vol. 26, No. 2, pp 71-130.
- [195] STONE M. (1959). Application of a measure of information to the design and comparison of regression experiments. *The annals of mathematical statistics*, Vol. 3, pp 55-70.
- [196] STREET, P. A. y STREET D. I. (1987). Combinations of experimental design. Oxford sciencie publications.
- [197] SZEGO G. (1939). Orthogonal Polynomials. American Mathematical Society providence. Rhode Island.
- [198] TSUKANOV A. V. (1981). On test of optimality of experimental designs for a regression model with random variables. *Theory Prob. Applications.*, Vol. 26, pp 173-177.
- [199] TITTERINGTON D.M. (1976). Algorithms for computing D-optimal designs on a finite design space. *Proceedings of the 1976 Conference on Information Sciences and Systems*.
- [200] TORSNEY B. (1981). A moment inequality and monotonicity of an algorithm. *Lecture Notes in Economics and Mathematical Systems. Semi-infinite Programming and Applications (Proceedings)*.
- [201] TORSNEY B. (1986). Moment inequalities via optimal design theory. *Linear Algebra and its Applications*, Vol. 82, pp 237-253.

- [202] TORSNEY B. (1988). Computing optimizing distributions with applications in design, estimation and image processing. *Optimal Design and Analysis of Experiments (proceedings)*. North Holland.
- [203] TORSNEY B. y ALAHMADI A. M. (1992). Further Development of Algorithms for Constructing Optimizing Distributions. *Proceedings of the 2nd HASA-workshop in St. Kyrik, Bulgaria*. Physica-Verlag, pp 121-129.
- [204] TORSNEY B. y ALAHMADI A. M. (1993). Designing for Minimally independent Observations. Preprint.
- [205] TOUTENBURG H. (1982). Prior information in linear models. Ed. John Wiley and Sons.
- [206] TSUKANOV A. V. (1981). On test of optimality of experimental designs for a regression model with random variables. *Theory Prob. Applications.*, Vol. 26, pp 173-177.
- [207] VUCHKOV I. N., DAMGALIEV D. L. y YONTCHEV CH. A. (1981). Sequencially generated second order quasi D-optimal designs for experiments with mixture and process variables. *Technometrics*, Vol. 23, No. 3, pp 233-238.
- [208] VUCHKOV I. N., YONTCHEV CH. A. y DAMGALIEV D. L. (1983). Continuous D-optimal designs for experiments with mixture and process variables. *Math. operationsforsch. statist.*, Ser. statistics, Vol. 14, No. 1, pp 33-51.
- [209] WACHTLER M. (1986). Cost-optimal performance of special reliability tests using prior knowledge. *Statistics* Vol.17, No. 11, pp 87-103.
- [210] WALD A. (1943). On the efficient design of statistical investigations. *The annals of mathematical statistics*, Vol. 14, pp 134-140.
- [211] WELCH W. J. (1982). Branch-and-bound search for experimental design based on D-optimality and other criteria. *Technometrics*, Vol. 24, No. 1, pp 41-48.
- [212] WHITE L. V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika*, Vol. 60, No. 2, pp 345-348.
- [213] WHITE L. V. y WELCH W. J. (1981). A method for constructing valid restricted randomization schemes using the theory of D-optimal design of experiments. *Journal royal statistical society, Ser. B*, Vol. 43, No. 2, pp 167-172.
- [214] WHITTLE P. (1973). Some general points in the theory of optimal experimental design. *Journal royal statistical society, Ser. B*, No. 1, pp 123-130.
- [215] WIHERICH W. (1985). Optimum design under experimental constrains for a covariate model and an intra-regression model. *Journal of statistical planing and inference*, Vol. 12, pp 27-40.
- [216] WIERICH W. (1986). On optimal designs an complete class theorems for experiments with continuous and discrete factors of influence. *Journal of statistical planning and inference*, Vol. 15, pp 19-27.
- [217] WU CH -F. (1978). Some algorithmic aspects of the theory of optimal designs. *The annals of statistics*, Vol. 6, No. 6, pp 1286-1301.

- [218] WU CH.-F. y WYNN H. P. (1978). The convergence of general step-length algorithms for regular optimum design criteria. *The annals of statistics*, Vol. 6, No. 6, pp 1273-1285.
- [219] WYNN H. P. (1970). The sequential generation of D-optimum experimental designs. *The annals of mathematical statistic*, Vol. 41, No. 5, pp 1655-1664.
- [220] WYNN H. P. (1972). Results in the theory and construction of D-optimum experimental design. *Journal royal statistical society, Ser. B*, Vol. B-34, pp 133-147.
- [221] WYNN H. P. (1975). Simple Conditions for Optimum Design Algorithms. J. N. Srivastava, ed., *A Survey of Statistical Design and Linear Models*. North-Holand Publishing Company, pp 571-579.
- [222] YATES F. (1970). *Experimental desing (Selected papers)*. Charles Griffin and Company Ltd.