

JUAN MANUEL RODRÍGUEZ DÍAZ

**CRITERIOS CARACTERÍSTICOS
EN DISEÑO ÓPTIMO DE EXPERIMENTOS**



EDICIONES UNIVERSIDAD DE SALAMANCA

COLECCIÓN VITOR

67

c

Ediciones Universidad de Salamanca
y Juan Manuel Rodríguez Díaz

1ª edición: Mayo, 2001
I.S.B.N.: 84 - 7800 - 871 - 3
Depósito Legal: S.670 - 2001
Ediciones Universidad de Salamanca
Apartado postal 325
E - 37080 Salamanca (España)

Edeltex S.L.
C/ Valle Inclán, 23 ,4ºB
37007 Salamanca
Tfno: 923 23 87 05

Impreso en España – Printed in Spain

*Todos los derechos reservados.
Ni la totalidad ni parte de este libro
puede reproducirse ni transmitirse
sin permiso escrito de
Ediciones Universidad de Salamanca*



CEP. Servicio de Bibliotecas

RODRÍGUEZ DÍAZ, Juan Manuel

Criterios Característicos en Diseño Óptimo de Experimentos

[Archivo de ordenador]

Juan Manuel Rodríguez Díaz – Salamanca :

Ediciones Universidad de Salamanca, 2001.

1 disco compacto. – (Colección Vitor; 67)

Tesis-Universidad de Salamanca, 2000

1. Diseño de experimentos

2. Universidad de Salamanca (España) - Tesis y disertaciones académicas.

519.242(043.2)

Resumen

El trabajo se sitúa en la teoría del Diseño Óptimo de Experimentos. Esta disciplina trata de realizar la mejor elección posible de los observables en los que se basará un experimento para obtener información acerca de un objeto; por ejemplo los mejores puntos sobre los que tomar observaciones para estimar los parámetros de un modelo de regresión. En los primeros capítulos se hace una revisión de conceptos básicos estadísticos y de esta teoría. Se introducen los principales criterios de optimización, que son los que permiten juzgar la bondad de un diseño. Muy importante es el Teorema de Equivalencia para comprobar cuándo un diseño es óptimo respecto de un criterio o cuándo distintos criterios son equivalentes. Se presentan asimismo los principales algoritmos para la obtención del diseño óptimo.

Todos estos conceptos y herramientas serán utilizadas posteriormente, en el momento de la definición de una nueva familia de criterios de optimización, los *Criterios Característicos*. Están basados en los coeficientes del polinomio característico de la matriz de información, matriz que caracteriza a cada diseño. Por tanto, la nueva familia de criterios incluye algunos de los más conocidos y utilizados: D - y A -optimización, que son esencialmente el determinante y la traza de la inversa de la matriz. Se estudian profundamente las funciones que darán lugar a estos criterios, obteniendo algunas propiedades destacables. Para los propios criterios se demuestran además ciertas de las cualidades más deseables: decrecimiento, convexidad y diferenciabilidad. Esta última se completa con la obtención explícita de su gradiente, el cual permitirá la aplicación de los algoritmos generales para la búsqueda de los óptimos característicos. Esto se realiza en el último capítulo para diferentes modelos, todos ellos ampliamente conocidos y utilizados en la práctica. Para el modelo polinómico en particular se propone un nuevo método de cálculo aproximado de óptimos característicos. Tanto los diseños calculados por este procedimiento como los obtenidos por los métodos algorítmicos son comparados sobre todo con los A - y D -óptimos. Resulta así que los óptimos para los nuevos criterios intermedios aparecen como una buena alternativa con que acercarse a la vez a varios de los objetivos tradicionales que persigue un buen diseño.

Para todos los cálculos ha sido necesaria la realización de un programa de ordenador, implementado con el programa *Mathematica*, cuyas características principales se describen en el Apéndice.

Es de destacar además la generalización de los nuevos criterios a una familia mayor que comprende además otro de los grupos más conocidos, los criterios Φ_p . Finalmente hay que mencionar las interpretaciones geométrica y estadística para los nuevos criterios, cuyos óptimos ofrecen una forma de minimizar las regiones de confianza de las estimaciones de subconjuntos de parámetros.

Summarize

The present work is set in the Optimal Experimental Designs theory. This is a statistics branch that tries to make the best election of the observable points. These are the ones in which an experiment is based on in order to get information about certain object. The Optimal Experimental Designs theory for instance looks for the best points to take observations in order to estimate the parameters of a regression model.

In the first chapters a revision of the basic statistical concepts and the theory is made. The main optimality criteria, allowing us to measure the goodness of a given design, are introduced. The Equivalence Theorem is also seen, to check when a design is optimal respect to a criterion and when different criteria are equivalent. The more known algorithms for obtaining optimal designs are revised as well.

The main objective is to define a new family of optimality criteria, called the *Characteristic Criteria*. In this subject a matrix is associated to each design, the information matrix of the design. And the new criteria are based on the characteristic polynomial of this matrix. Thus, the new family includes some of the most known and used criteria: D - and A -optimality, essentially the determinant and the trace of the inverse of the information matrix. The functions that are used to define these new criteria are deeply studied, obtaining some relevant results. For these criteria some of the most important properties are proved, e.g. that they are decreasing for the loewner ordering and convex. The differentiability is seen as well and the gradient computed, what allow us to apply the general algorithms for searching optimal characteristic designs.

The last chapter is due to this activity, obtaining optimal characteristic designs for different known and widely used models. For the polynomial model is developed a new method for obtaining approximated support points for characteristic optimal designs.

All the designs computed are compared with the popular A - and D -optimal designs. The new ones seem to be a good alternative when we look for several of the desirable objectives that are aimed in a design.

A computer implementation has been made in order to carry out all the calculus. It has been developed with the program *Mathematica*, and its basic structure is shown in the Appendix.

Finally it is worthwhile to note that the new criteria have been generalized to a biparametric set of optimality criteria that includes also the known family of Φ_p criteria. The geometric and statistical meaning of the optimal characteristic designs are studied as well. They provide a way to minimize the confidence regions for the joined estimation of subsets of parameters.

Sumario

Introducción	1
1 Modelos Lineales	5
1.1 Nociones generales	6
1.2 Método de los mínimos cuadrados generalizados	9
1.3 Regresión	10
1.3.1 Estimación de parámetros	12
1.3.2 Principales modelos	13
1.4 Diseños clásicos de ANOVA	15
1.5 Introducción a los modelos no lineales	17
2 Diseño Óptimo de Experimentos	19
2.1 Panorama histórico	20
2.2 Etapas	24
2.3 Conceptos básicos	25
2.3.1 Diseños exacto y aproximado	27
2.3.2 Estimadores de funcionales lineales	28
2.4 Criterios de optimización	38
2.4.1 Introducción	38
2.4.2 Funciones criterio	40
2.5 Teorema de equivalencia	56
2.6 Crítica al Diseño Óptimo	62
3 Algoritmos	66
3.1 Algunos algoritmos conocidos	66
3.1.1 Aproximación al diseño óptimo	66
3.1.2 Procesos iterativos	68
3.2 Diseños sin replicación para conjuntos finitos	72

3.2.1	Cálculo del <i>diseño sin replicación</i> óptimo	72
3.2.2	Cálculo de un buen <i>diseño sin replicación</i>	73
3.3	Algoritmo Sistemático	75
3.3.1	Algoritmo local	75
3.3.2	Algoritmo Sistemático	77
4	Órdenes de las matrices de información	83
4.1	Orden de Loewner	83
4.2	Propiedades	84
4.2.1	Matrices <i>NND</i> y orden de Loewner	84
4.2.2	Relación “ \prec ”	85
4.2.3	Relación “ \preceq ”	87
5	Criterios característicos	91
5.1	Introducción y definiciones	91
5.1.1	Funciones simétricas elementales	91
5.1.2	Propiedades	92
5.1.3	Funciones características	97
5.2	Continuidad y Diferenciabilidad	99
5.3	La segunda función característica	104
5.4	Interpretación geométrica	108
6	Cálculo de óptimos característicos	110
6.1	Descripción del algoritmo general	110
6.2	Modelo polinómico	111
6.2.1	Eficiencias	112
6.2.2	Relaciones	115
6.2.3	Cálculo aproximado de los Φ_{Ch_k} -óptimos	116
6.3	Otros modelos	120
6.3.1	Modelo exponencial	120
6.3.2	Modelo compartimental	121
	Conclusiones	129
A	Programa CharCriteria	128
	Bibliografía	136
	Índice de materias	150

Lista de Figuras

2.1	Método gráfico debido a Elfving para calcular el diseño c -óptimo para el modelo biparamétrico	53
5.1	Interpretación geométrica del Φ_{Ch_2} -óptimo para el modelo cuadrático: minimiza el promedio de las áreas de las regiones de confianza de la estimación conjunta de los parámetros tomados de dos en dos	109
6.1	Representación de los Φ_{Ch_k} -óptimos para los modelos de grados 3 y 4, con k variando entre 1 y 7	115
6.2	Representación de los Φ_{Ch_2} y Φ_{Ch_3} -óptimos para los modelos de grados 3 a 7	116
6.3	“Suave” disminución en los pesos de los puntos del soporte del A -óptimo al representar los diseños óptimos para los modelos polinómicos de grado n creciente	117
6.4	Relación entre los A -óptimos para los modelos polinómicos al representarlos paralelamente según varía el grado del modelo ($n = 2, \dots, 12$)	118

Lista de Tablas

6.1	Diseños óptimos característicos para los modelos polinómicos de grados 2 a 6	113
6.2	Eficiencias de los diseños óptimos característicos para los modelos polinómicos de grados 2 a 6	114
6.3	A -óptimos para los modelos polinómicos de grados 2 a 12 . . .	119
6.4	A -eficiencias (%) de varios diseños con soporte aproximado y pesos A -óptimos para ese soporte. Se incluyen los soportes generados con el procedimiento descrito en el Capítulo 6, así como el soporte arcoseno	120
6.5	D -óptimos para los modelos polinómicos de grados 2 a 12 . . .	121
6.6	D -eficiencias (%) de varios diseños con soporte aproximado y pesos iguales. Se incluyen los soportes generados con el procedimiento descrito en el Capítulo 6, así como el soporte arcoseno	122
6.7	Diseños óptimos característicos para el modelo exponencial . .	122
6.8	Eficiencias de los diseños óptimos característicos para el modelo exponencial	123
6.9	Diseños óptimos característicos para el modelo compartimental en el intervalo (0,50)	124
6.10	Eficiencias de los diseños óptimos característicos para el modelo compartimental en el intervalo (0,50)	124

Introducción

En la mayor parte de las disciplinas, tras un estudio detallado de un cierto tema, se intenta generalizar las conclusiones obtenidas, y en muchos casos esto pasa por la formulación de un modelo matemático que refleje adecuadamente estos resultados. Esta fórmula envolverá las variables objeto de estudio con parámetros que sirven de “enlace” entre ellas para expresar el resultado final. En un principio estos parámetros son desconocidos, pero se puede intentar extraer información sobre ellos a partir de la observación de una serie de pruebas, esto es, mediante la realización de un experimento. Podríamos definir *experimento científico* como un conjunto, más o menos complejo, de actividades realizadas con el fin de alcanzar un conocimiento profundo acerca de un objeto.

Un experimento, pues, 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.

De lo dicho anteriormente se desprende la necesidad de optimizar los resultados finales empleando para ello la menor cantidad posible de observaciones experimentales. 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.

El presente trabajo se basa en esta teoría. El primer capítulo aborda una revisión general de los modelos lineales, en los que se basa fundamentalmente la teoría del Diseño Óptimo de Experimentos. Se introduce la notación que será seguida a lo largo de toda la exposición y ciertas nociones y métodos

generales de estimación estadística y regresión. Se presentan los modelos lineales más utilizados, incluyendo los modelos de ANOVA. Al final hay una breve referencia a los modelos no lineales, cuyo tratamiento es, como puede apreciarse, mucho más complicado.

El siguiente capítulo se centra en el tema del diseño óptimo de experimentos, teoría general en la que está basado todo el conjunto. La primera parte incluye una breve introducción histórica así como una descripción de los principales conceptos y etapas relativos al diseño óptimo. Se introducen los conceptos de diseño exacto y aproximado que serán utilizados en los siguientes capítulos y se especifican los objetivos generales en la búsqueda del diseño óptimo. A partir de entonces se inicia la definición de los principales criterios de optimización conocidos y utilizados, así como su interpretación estadística y principales propiedades y relaciones entre ellos. El Teorema de Equivalencia, uno de los principales pilares de toda la teoría, identifica los diseños óptimos aproximados respecto de diferentes criterios. Proporciona un instrumento inestimable para la comprobación de la optimalidad de un diseño. No requiere el uso de criterios diferenciables y facilita algoritmos de construcción de diseños óptimos. En muchos casos ofrece un método rápido de cálculo de un óptimo cuando se tienen fundadas sospechas sobre la forma del diseño óptimo. La última sección expone las críticas principales realizadas sobre estos procedimientos, así como una defensa razonada de las mismas.

El Capítulo 3 presenta una relación de los principales algoritmos existentes en la literatura para el cálculo explícito del diseño óptimo respecto de funciones criterio generales y particulares. Entre ellos destaca el de Fedorov-Wynn para funciones criterio convexas, que será utilizado en capítulos posteriores. Se incluyen además algunos algoritmos originales del autor para la búsqueda del óptimo cuando el conjunto soporte es discreto.

Los algoritmos buscarán el óptimo a partir de uno o varios diseños iniciales, modificándolos en cada etapa en la dirección adecuada. Surge entonces el problema de la elección de estos diseños iniciales y también de los que han de ir incorporándose en cada paso. Parece conveniente definir un orden en el conjunto de todos los diseños que nos permita escoger los más adecuados. Cada diseño tiene asociada una matriz simétrica semidefinido positiva cuya inversa es proporcional a la matriz de covarianzas de los estimadores de los parámetros. Un buen diseño, por tanto, buscará “minimizar” de alguna manera esta matriz. Esta idea provocó la necesidad del desarrollo del siguiente capítulo, en el que se estudia la ordenación de los diseños a través de sus ma-

trices de información. Es el momento de presentar una serie de propiedades novedosas, basadas principalmente en el orden de Loewner de las matrices semidefinido positivas. Se estudian principalmente dos relaciones de orden. La investigación se centra sobre todo en los diseños unipuntuales y sus combinaciones convexas porque se identifican naturalmente con los puntos que han de ser incorporados en cada paso de los algoritmos.

Seguidamente, en el Capítulo 5 se aborda la tarea de la creación de nuevos criterios basados en los coeficientes del polinomio característico de una matriz. Esto provocará la generalización de algunos de los criterios más conocidos y utilizados hasta el momento, como son A - y D -optimización. Estos criterios se pueden obtener tomando el valor absoluto de los coeficientes del polinomio característico de la inversa de la matriz de información. Como se verá, es posible modificar esta definición inicial con el fin de buscar propiedades deseables, como la homogeneidad, o incluso para crear una familia general de criterios que incluya a la conocida Φ_p . Sin embargo, la primera forma será la más utilizada debido a su sencillez. El capítulo introduce por tanto una teoría consistente y original debido a lo novedoso del tema. Primeramente se abordarán las definiciones, notación y propiedades generales, para después concentrarse en casos más específicos. Se prestará especial atención al criterio originado por el segundo coeficiente, por ser el más sencillo de los nuevos. Este capítulo es la parte central del volumen, y la que dará lugar a la serie de aplicaciones prácticas que aparecen en el Capítulo 7.

En éste se calculan explícitamente los diseños óptimos respecto de los nuevos criterios para distintos modelos, mediante el desarrollo e implementación de un programa informático. Se estudia especialmente el modelo polinómico debido a su popularidad y utilidad, aunque se trabaja brevemente sobre otros, lineales y no lineales. Esto nos permitirá presentar una comparación objetiva de nuestros criterios con los ya existentes para los modelos más habituales, mediante el cálculo de las eficiencias de los diseños obtenidos. Se presenta a su vez un procedimiento original para el cálculo de los óptimos aproximados respecto de los nuevos criterios para el modelo polinómico. Este método, como puede comprobarse a tenor de los resultados obtenidos, proporciona unos diseños altamente eficientes. Se muestran además tablas en las que se comparan los diseños así construídos con otros conocidos que utilizan también soportes aproximados.

El último capítulo recoge las conclusiones, en las que se enumeran las aportaciones originales que incorpora este trabajo.

El Apéndice A contiene una descripción elemental del programa utilizado para el cálculo de los óptimos característicos definidos en el Capítulo 5, que ha sido implementado con el programa *Mathematica*. Con el fin de evitar la aparición de múltiples páginas de engorrosos listados se describen someramente las principales funciones que incorpora el programa, generales y particulares, sin entrar en detalles de programación.

Al final aparece la bibliografía. Además de las obras citadas a lo largo del volumen se incluyen otras de propósito general sobre el Diseño Óptimo, así como referencias obligadas que han marcado hito en la historia de esta materia.

Capítulo 1

Modelos Lineales

La Estadística actual es la unión del Cálculo de Probabilidades, disciplina surgida con el fin de estudiar los juegos de azar, y de la Estadística propiamente dicha, la ciencia de la descripción de datos. Actúa como puente entre los modelos matemáticos y los fenómenos reales, analizando las diferencias que proporcionan los datos experimentales con los teóricos que se supondrían a partir del modelo. Es la base matemática última en la que fundamentar los resultados finales que obtienen las ciencias experimentales (química, sociología, medicina,...), y como tal su importancia no ha dejado de crecer paulatinamente, hasta hacerse casi indispensable en la mayoría de las disciplinas actuales.

Mediante el experimento científico obtenemos unas observaciones que nos van a permitir ajustar un modelo a la realidad que es objeto de estudio. Dicho modelo debe contener una descripción del estado del objeto observado. En principio consideraremos modelos lineales en los que se conoce de antemano el conjunto de puntos observables, llamado *espacio del diseño*. Los estados de la naturaleza serán las posibles realizaciones del modelo que representarán a la realidad. Las observaciones hechas en los puntos que el diseño haya seleccionado serán consideradas como variables aleatorias.

El Diseño Óptimo de Experimentos está íntimamente ligado a los modelos lineales, es decir, al análisis de la varianza, la regresión y el análisis de la covarianza. El objetivo es lograr la mejor estimación posible de los parámetros desconocidos que aparecen en el modelo. Nuestro cometido será elegir adecuadamente los puntos sobre los que realizar la prueba para que dichos estimadores tengan la menor varianza posible, tratando a la vez de

minimizar las covarianzas. La manera en que se va a hacer esto se verá más adelante.

A continuación desarrollaremos brevemente una serie de conceptos básicos y generales en estadística, fundamentalmente con el objetivo de introducir una notación estándar que usaremos en todo este trabajo.

1.1 Nociones generales

El objetivo principal de la estadística es intentar obtener conclusiones acerca de diversos aspectos característicos de la *población* objeto de estudio, a partir de la observación de unos cuantos valores Y_1, \dots, Y_N de la misma extraídos al azar, a los que se llama *muestra*. Las deducciones podrán en general variar ampliamente debido al hecho de que la muestra es aleatoria, y por tanto los posibles instrumentos de decisión que construyamos a partir de ella nunca tomarán valores determinados totalmente.

En muchos casos se posee una gran evidencia sobre la distribución de probabilidad de la población (que es una representación de la frecuencia real con que la población toma cada uno de los distintos valores que pueda llegar a valer). Entonces reduciríamos el conocimiento de la población simplemente al de ciertos parámetros que caracterizan a la misma (por ejemplo, su media y varianza) y que necesitamos para definir completamente su función de distribución. Éstos se suelen denotar con letras griegas (media= μ , varianza= σ^2). Un problema de estadística será por tanto un problema en el cual se han de analizar datos que han sido generados de acuerdo con una distribución de probabilidad desconocida, y en el que se debe realizar algún tipo de inferencia acerca de tal distribución (que podría ser especificarla totalmente), de sus propiedades, de su relación con otras, etc. Cuando hay un número suficiente de observaciones vamos a poder suponer en muchos casos que la distribución de los estadísticos que usamos es normal, o una de las que se derivan de ella.

Los parámetros que se intentan determinar son principalmente la *media* y la *varianza*. Se denomina *Esperanza* o *Media* de una cierta variable aleatoria X al promedio de los valores que puede tomar la variable. Esto es, la suma (en el caso discreto) o integral (caso continuo) de los posibles valores x de la variable multiplicados cada uno por la probabilidad o densidad de la variable

en el punto x . Se denota $\mu_X = E[X]$. La *varianza* de X sería la esperanza de la función que mide las diferencias cuadráticas de cada uno de los valores de X con su media, esto es $E[(X - \mu)^2]$. Se suele denotar σ_X^2 . La *covarianza* entre dos variables X e Y se define como $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$ y mide la relación lineal existente entre las variables. Dado el hecho de que ésta depende de las unidades de medida, se suele emplear el *coeficiente de correlación lineal*, definido como

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

En algunos problemas de estadística se tiene cierto control sobre el lugar y la proporción de datos experimentales que se van a recoger. Estos problemas, en los que el experimentador puede elegir, al menos hasta cierto punto, el experimento concreto que se va a llevar a cabo, se llaman problemas de *diseño de experimentos*. Por supuesto, el diseño de un experimento y el análisis estadístico de los datos experimentales están estrechamente relacionados: para diseñar adecuadamente un experimento conviene tener en cuenta el análisis estadístico que se realizará con los datos que se van a obtener; y no se debería llevar a cabo un análisis estadístico de datos experimentales sin considerar el tipo concreto de experimento del cual se obtienen los datos.

Centrémonos ahora en el tema de inferencia estadística en el que se intentará asignar un valor al parámetro (o vector paramétrico) desconocido θ que deberá pertenecer a algún espacio paramétrico Ω . Su estimación dependerá de la muestra observada, siendo una función de la misma. Cada uno de los valores que puede tomar esta función sería una *estimación* del parámetro.

En muchos problemas, antes de observar la muestra X_1, \dots, X_n , el experimentador podrá resumir su información y conocimientos previos acerca de donde es probable que se encuentre el valor de θ dentro del espacio paramétrico Ω , construyendo así una distribución de probabilidad para θ en el conjunto Ω . Esta distribución se denomina *distribución inicial* o *a priori* de θ , porque representa nuestra información sobre θ *antes* de observar la muestra. Luego, a partir de ésta, podemos llegar a conocer la *distribución final* o *a posteriori* de θ , para la cual ya utilizamos toda la información que nos proporciona la observación de X_1, \dots, X_n . Finalmente, tras elegir una *función de pérdida* adecuada, que refleje el coste que supondrá para el decisor el fallo de la predicción, tomaríamos como estimador θ^* del parámetro θ aquel valor

que minimizara la pérdida esperada de la distribución final. El estimador así obtenido se llama *estimador Bayes* de θ .

Según el procedimiento explicado, para hallar el estimador Bayes de un parámetro desconocido θ es necesario primeramente partir de una distribución inicial y además fijar una función de pérdida adecuada. La elección de ambas es, en muchas ocasiones, bastante subjetiva, dependiendo del conocimiento que sobre el tema posee el experimentador, pero también de su experiencia, intuición y otros factores difícilmente cuantificables, por lo que la elección de este método de estimación así como la de la distribución inicial y la función de pérdida, resulta en muchas ocasiones un tema controvertido. Los estadísticos que piensan que siempre es conviene asignar una distribución inicial al parámetro se dice que se adhieren a una filosofía bayesiana de la estadística.

No todo el mundo comparte este punto de vista, y así surgen otros métodos de estimación igualmente válidos y ampliamente aceptados. Por ejemplo la *estimación máximo verosímil*, mediante la que, tras observar la muestra, que se supone procede de una familia paramétrica de distribuciones, se elige como valor más probable del parámetro θ el que maximiza la probabilidad de que ocurra lo que ya ha ocurrido. Es decir, tomaríamos como estimador, $\hat{\theta}$, el que hiciera más verosímil la obtención de la muestra X_1, \dots, X_N de entre todos los posibles valores de θ . Por tanto es el que maximiza la función de verosimilitud (función de densidad conjunta tomada como función del parámetro) de X_1, \dots, X_N .

En cualquier caso, diremos que $\hat{\theta}$ es un *estimador centrado* del parámetro desconocido cuando su esperanza sea el propio parámetro. La varianza del estimador será entonces el error cuadrático medio de dicha estimación.

En distintas ocasiones nos encontraremos ante el hecho de intentar expresar una variable Y en función de otra u otras X_1, \dots, X_m (véase (1.3)). Esto se podría escribir de la forma

$$Y = \eta(X, \beta) + \varepsilon,$$

donde β representa un conjunto de parámetros desconocidos que determinan completamente la función η , llamada *modelo de regresión*. La hipótesis habitual es que se verifica $E[\varepsilon] = 0$, por lo que una manera alternativa de escribir el modelo es

$$E[Y] = \eta(X, \beta).$$

Nos interesará especialmente el *modelo lineal en los parámetros*, con la hipótesis adicional de que las observaciones son independientes.

1.2 Método de los mínimos cuadrados generalizados

La identidad del descubridor del método de los mínimos cuadrados sigue siendo un tema controvertido. Parece que fue descubierto independientemente por Carl Friedrich Gauss (1777-1855) y Adrien Marie Legendre (1752-1833). Se cree que Gauss comenzó a usarlo antes de 1803 (según él sobre 1795, pero no se ha podido corroborar este dato). Apareció publicado por primera vez por Legendre en 1805. Cuando Gauss escribió en 1809 que él había usado este método antes de la fecha de la publicación de Legendre, comenzó la polémica. Una de sus primeras aplicaciones fue en el cálculo de órbitas de planetas.

El método tal y como es usado hoy en día en estadística es el siguiente:

Consideremos un modelo general

$$Y = \eta(X, \beta) + \varepsilon$$

siendo Σ_ε la matriz de covarianzas de los errores.

La suma de cuadrados de los errores ponderada por la matriz de covarianzas de las observaciones es entonces

$$\varepsilon^t \Sigma_\varepsilon^{-1} \varepsilon = [Y - \eta(X, \beta)]^t \Sigma_\varepsilon^{-1} [Y - \eta(X, \beta)].$$

El estimador mínimo cuadrático de β es el valor $\hat{\beta}$ que al ser sustituido en la ecuación anterior minimiza $\varepsilon^t \Sigma_\varepsilon^{-1} \varepsilon$. Habitualmente se puede calcular derivando la ecuación respecto de β e igualando a cero.

En el caso lineal, detallado en 1.3.1, la solución tiene las siguientes propiedades:

- Es un estimador de β que minimiza la suma generalizada de cuadrados de los errores $\varepsilon^t \Sigma_\varepsilon^{-1} \varepsilon$, sean cuales sean las propiedades de la distribución de éstos.

- Los elementos de $\hat{\beta}$ son funciones lineales de las observaciones Y_1, \dots, Y_N , y proporcionan estimadores centrados de β con varianza mínima entre todas las funciones lineales de las observaciones que proporcionan estimadores centrados, sea cual sea la distribución de los errores.
- Si los errores se distribuyen normalmente con media 0 y varianza constante σ^2 , entonces $\hat{\beta}$ es el estimador máximo verosímil de β . Esto es debido a que la función de verosimilitud para la muestra sería en este caso

$$f(Y_1, \dots, Y_N) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \varepsilon^t \varepsilon \right\},$$

así que para un valor fijo de σ , maximizar la función de verosimilitud equivale a minimizar $\varepsilon^t \Sigma_\varepsilon^{-1} \varepsilon$. Esto es una buena justificación del empleo del procedimiento de mínimos cuadrados.

En los modelos no lineales suele ser necesario recurrir a algoritmos numéricos, como el de Marquart.

Si hemos utilizado el método de mínimos cuadrados para estimar β por $\hat{\beta}$, estén los errores distribuidos normalmente o no, tenemos los siguientes resultados:

- El vector de residuales es $e = Y - \hat{Y}$
- En el caso del modelo lineal con varianza constante σ^2 , la matriz de covarianzas de los estimadores de los parámetros es $\Sigma_{\hat{\beta}} = (X^t X)^{-1} \sigma^2$. La varianza σ^2 se puede estimar por máxima verosimilitud, y desde el punto de vista del diseño óptimo de experimentos se puede eliminar sin más que introducirla en el modelo, aunque no sea constante.

1.3 Regresión

La teoría de la regresión consiste en la construcción de modelos estadísticos para representar la dependencia de una variable respuesta continua respecto de una o más variables explicativas también continuas. El nombre proviene de los estudios de Galton sobre la dependencia entre la estatura de los hijos y la estatura de los padres, encontrando que las estaturas de los hijos de

padres altos iban disminuyendo en promedio, lo contrario que ocurría con las de los hijos de padres bajos, y que por tanto se producía una “regresión” a la media.

Llamaremos y a la variable dependiente, y x_1, \dots, x_m al conjunto de variables explicativas, y nos centraremos en el *Modelo de Regresión Lineal*

Cada una de las posibles realizaciones del modelo, es decir, los posibles valores de los parámetros, se llama *estado*, θ , y el conjunto de todos los posibles estados se denota por Θ .

Cuando tomamos N observaciones el modelo se puede escribir matricialmente:

$$Y = X\beta + \varepsilon,$$

donde

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix},$$

siendo Y el vector de observaciones y ε el vector de los errores muestrales. La matriz del diseño vendrá dada por

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nm} \end{pmatrix}$$

Estamos suponiendo $E[\varepsilon] = 0$, con matriz de covarianzas

$$\Sigma_\varepsilon = E[\varepsilon \varepsilon^t] = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1N}^2 \\ \sigma_{12}^2 & \sigma_2^2 & \cdots & \sigma_{2N}^2 \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{1N}^2 & \sigma_{2N}^2 & \cdots & \sigma_N^2 \end{pmatrix}$$

que es el caso más general con heterocedasticidad y observaciones correladas. Además no suponemos de momento normalidad (lo que identificaría los estimadores máximo verosímiles y mínimo cuadráticos).

1.3.1 Estimación de parámetros

Veamos cómo realizar la estimación de los parámetros:

Supondremos que Σ_ε es una matriz conocida (por ejemplo se estimaría con datos retrospectivos). Aplicamos el método de mínimos cuadrados (1.2), que consiste en minimizar el promedio del cuadrado de las diferencias entre las observaciones y sus estimaciones, convenientemente ponderado por la matriz de covarianzas:

$$Q(\beta) = (Y - X\beta)^t \Sigma_\varepsilon^{-1} (Y - X\beta)$$

y queremos minimizar esta función, para lo que habremos de derivar la expresión

$$(Y - X\beta)^t \Sigma_\varepsilon^{-1} (Y - X\beta) = Y^t \Sigma_\varepsilon^{-1} Y - 2\beta^t X^t \Sigma_\varepsilon^{-1} Y + \beta^t X^t \Sigma_\varepsilon^{-1} X \beta.$$

Teniendo en cuenta que

$$\nabla(\beta^t A \beta) = \frac{\partial(\beta^t A \beta)}{\partial \beta} = 2A\beta; \quad \nabla(a^t \beta) = \frac{\partial(a^t \beta)}{\partial \beta} = a;$$

$$\nabla(A\beta) = \frac{\partial(A\beta)}{\partial \beta} = A^t$$

obtenemos

$$0 = -2 X^t \Sigma_\varepsilon^{-1} Y + 2 X^t \Sigma_\varepsilon^{-1} X \beta.$$

Podrían darse dos casos: o bien tenemos m ecuaciones independientes con m incógnitas (los elementos de $\hat{\beta}$), o algunas de las ecuaciones dependen de las otras. En este último caso $X^t \Sigma_\varepsilon^{-1} X$ es singular, esto es, no es de rango máximo, luego no existe su inversa. Entonces el modelo debería ser expresado con menor número de parámetros, o se deberían suponer restricciones adicionales sobre los mismos, ya que si no las estimaciones de los parámetros se confundirían entre sí, dependiendo del diseño empleado. Para obtener los estimadores habría que recurrir entonces a la elección de una inversa generalizada. En el caso de que las m ecuaciones normales fueran independientes, $X^t \Sigma_\varepsilon^{-1} X$ sería regular y existiría su inversa. La solución de las ecuaciones normales podría escribirse

$$\hat{\beta} = (X^t \Sigma_\varepsilon^{-1} X)^{-1} X^t \Sigma_\varepsilon^{-1} Y \quad (1.1)$$

lo que equivale a transformar las variables para que $\Sigma_\varepsilon = Id \sigma^2$ y hallar el estimador mínimo cuadrático ordinario. Puesto que $\Sigma_{AY} = A\Sigma_Y A^t$ se tiene

$$\Sigma_{\hat{\beta}} = \left(X^t \Sigma_\varepsilon^{-1} X\right)^{-1} X^t \Sigma_\varepsilon^{-1} \Sigma_Y \Sigma_\varepsilon^{-1} X \left(X^t \Sigma_\varepsilon^{-1} X\right)^{-1} = \left(X^t \Sigma_\varepsilon^{-1} X\right)^{-1}$$

Como veremos más tarde, se define $M = X^t \Sigma_\varepsilon^{-1} X$ como la matriz de información. Además

$$E[\hat{\beta}] = \left(X^t \Sigma_\varepsilon^{-1} X\right)^{-1} X^t \Sigma_\varepsilon^{-1} X \beta = \beta$$

lo que prueba que $\hat{\beta}$ es un estimador centrado de β .

Obsérvese que bajo la hipótesis de normalidad, la función de densidad sería

$$f(Y|X\beta, \Sigma_\varepsilon) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma_\varepsilon}} \exp \left[-\frac{1}{2} (Y - X\beta)^t \Sigma_\varepsilon^{-1} (Y - X\beta) \right],$$

de modo que en este caso coincidirían los estimadores máximo verosímiles y mínimo cuadráticos como ya se observó en el caso homocedástico de observaciones independientes.

1.3.2 Principales modelos

Establezcamos ahora algunos tipos de modelos de uso frecuente en la literatura:

- Modelo de observaciones directas: $\Theta = \{\alpha_j : \alpha \in \mathbb{R}, \text{ donde } j(x) = 1, x \in \Xi\}$. 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: $\Theta = \{\sum_{i=1}^m \alpha_i f_i : (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m\}$, donde f_1, \dots, f_m son funciones continuas y linealmente independientes. Todo modelo lineal con observaciones incorreladas puede reducirse a éste.

- Modelo con emparejamiento lineal: X finito,

$$\Theta = \left\{ \theta \in \mathcal{C}(X) : \sum_{x \in \Xi} b_i(x)\theta(x) = 0 \text{ para } i = 1, \dots, q \right\} \quad (1.2)$$

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 (1.2) 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 con restricciones lineales sobre los parámetros:

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

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 disminuyendo el número de parámetros a $m - q$ mediante la resolución del sistema de ecuaciones que forma la restricción.

- 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 $f_1^{(i)}, \dots, f_{m_i}^{(i)}$ son funciones linealmente independientes en el intervalo $[x_{i-1}, x_i]$, $i = 1, \dots, k$. El espacio de estados será:

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

Es un modelo de regresión a trozos que conecta con suavidad en los extremos de los intervalos de la partición. 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 $j=1$. Para $r = 1$ se tendrán las siguientes condiciones de conexión:

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

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

y habría que estimar solamente dos parámetros.

1.4 Diseños clásicos de ANOVA

El objetivo de un experimento es estudiar el efecto que sobre una variable de interés, *variable respuesta*, tienen un conjunto de otras variables, que llamaremos *variables explicativas* o *factores*. En general supondremos que la variable respuesta es continua y que los factores se fijan antes del experimento a ciertos niveles determinados, es decir, tienen efectos fijos. El experimento consiste en fijar los valores de los factores a distintos niveles y observar el valor de la respuesta. El tamaño del experimento será el número total de respuestas obtenidas, es decir, el número de experimentos llevados a cabo.

La metodología del diseño de experimentos estudia cómo realizar comparaciones lo más homogéneas posibles, para aumentar la probabilidad de detectar cambios o identificar variables influyentes. Solamente podemos confiar en detectar relaciones de causalidad mediante un experimento bien diseñado.

En cualquier experimento existen a priori un gran número de variables que pueden influir sobre los resultados. Algunas serán de interés, pero otras no, y por tanto convendría eliminar su efecto. Para ello se tiene tres caminos posibles: mantenerlas fijas durante la realización del experimento, hacer que las comparaciones de interés se realicen para valores fijos de estas variables, o evitar su influencia aleatorizando su aparición en los diversos tratamientos.

Los dos primeros métodos se emplean para el caso de variables controladas por el experimentador, mientras que la aleatorización se reserva para eliminar el efecto de variables fuera de nuestro control y de poca influencia esperada, cuyos efectos se englobarán dentro del error experimental.

Con el fin de lograr la aleatorización se desarrollan distintos diseños. En cada uno de ellos se intenta combinar lo máximo posible los niveles de los distintos factores a la hora de tomar las observaciones. Los principales son:

- Modelo en bloques aleatorizados: Existe un tratamiento de interés, mientras que el resto de variables (llamadas bloques) se suponen independientes entre ellas y con el tratamiento, y sólo se introducen en el modelo para lograr una mayor aleatorización. El modelo más sencillo sería

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

donde α es el efecto debido al tratamiento y β el que se deriva de la existencia de bloques.

- Modelo factorial con varios factores e interacciones: Hay varios tratamientos de interés, y estos podrían interactuar entre sí, por lo que el modelo completo resultaría

$$y_{ijk\dots} = \mu + \alpha_i + \beta_j + \gamma_k + \dots + (\alpha\beta)_{ij} + \dots + (\alpha\beta\gamma)_{ijk} + \dots + \varepsilon_{ijk\dots}$$

donde $\alpha, \beta, \gamma, \dots$ son los factores y $(\alpha\beta\dots)_{ijk\dots}$ las interacciones. Podrían considerarse modelos incompletos. De hecho, las interacciones mayores que la tercera no se suelen incluir.

- Cuadrado Latino: Se puede aplicar cuando existen tres factores con el mismo número de niveles cada uno, I , y entre los que no se espera la existencia de interacciones. Sólo habría que tomar I^2 observaciones, pero teniendo en cuenta que cada uno de los niveles de cada factor tiene que aparecer con cada nivel de los demás factores en alguna observación. El modelo sería pues

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

donde α sería el *efecto fila*, β el *efecto columna* y γ el *efecto letra*.

- Cuadrado Grecolatino: Basándose en la misma idea que el anterior, este se puede usar cuando existen cuatro factores con el mismo número de niveles cada uno, I , y entre los que no se espera la existencia de interacciones. De manera similar al cuadrado latino, habría que tomar I^2 observaciones, pero teniendo en cuenta que cada uno de los niveles de cada factor tiene que aparecer con cada nivel de los demás factores en alguna observación. En este caso el modelo sería

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl}$$

donde α sería el *efecto fila*, β el *efecto columna*, γ el *efecto letra latina* y δ el *efecto letra griega*.

- Modelo con efectos aleatorios: A diferencia de los efectos fijos, en este caso los niveles de los factores no están prefijados de antemano, no los elige el experimentador, sino que más bien se determinan de manera aleatoria dentro de cierta población. Los modelos de efectos fijos se utilizan para conocer si deben o no aplicarse determinados tratamientos, mientras que los modelos de efectos aleatorios permiten medir el efecto de factores que están presentes en el proceso, sean o no de interés. Los efectos fijos son susceptibles de ser tratados desde el punto de vista del diseño óptimo.

1.5 Introducción a los modelos no lineales

Cuando el modelo de regresión es no lineal en los parámetros no se pueden emplear los procedimientos descritos anteriormente. El tratamiento de estos modelos conlleva una mayor dificultad. Una manera de enfrentarse a ellos es aproximada, *linealizándolos*, esto es expresar el modelo mediante su desarrollo como serie de Taylor y quedarse con la parte lineal. En el método general se utiliza la *matriz de información de Fisher* como herramienta para la estimación de los parámetros, cuando se asume normalidad en las observaciones. El inconveniente que se presenta es que esta matriz depende de los propios parámetros, por lo que necesitaremos unos valores iniciales de los mismos para obtener las estimaciones deseadas. Dicho de otro modo, los diseños óptimos (aquellos que proporcionan una mejor estimación de los parámetros,

y que serán ampliamente tratados en capítulos sucesivos) conseguidos mediante este procedimiento van a depender de los propios parámetros, y para cada valor inicial de éstos proporcionarán distintos diseños. Se trata de los llamados *diseños localmente óptimos*, es decir, óptimos cuando se sabe que los parámetros están muy cercanos a ciertos valores específicos. En este sentido puede consultarse Chernoff (1953) o Frischmuth (1989). Otro método tradicional es transformar el modelo cuando sea posible, para hacerlo lineal en otros parámetros.

El trabajo de Chaloner y Verdinelli (1995) es una revisión bastante completa de los métodos bayesianos aplicados tanto a la regresión lineal como no lineal, para estimación y predicción. En él tratan la búsqueda del mejor diseño como un problema de decisión, eligiendo la *función de utilidad* adecuada al propósito del experimento, y buscando el diseño que maximice la utilidad esperada. Se distingue entre diseños óptimos para estimación o para predicción, y se identifica algunos de los *criterios alfabéticos* ampliamente utilizados (véase cap. 2) como casos particulares de los métodos bayesianos, eligiendo adecuadamente la función de utilidad. El abanico de elección de ésta proporciona una amplia gama de aplicaciones, además de criterios, para estos métodos, adecuados para muy diversos objetivos.

Capítulo 2

Diseño Óptimo de Experimentos

Los modelos matemáticos son empleados cada vez con más frecuencia en un amplio abanico de disciplinas experimentales, que se va incrementando paulatinamente desde el momento en que en todas ellas se intenta expresar de una manera exacta y objetiva los resultados obtenidos. Se hace necesaria, por tanto, una buena organización de la manera en la que se va a realizar todo el proceso, las pruebas que se van a realizar, la recogida de los datos, etc. Entra aquí en juego el *diseño de experimentos* que, basándose en el principio de aleatorización (en general no podremos controlar todo el proceso), intenta eliminar el efecto de las variables extrañas para tener en cuenta solamente las que de verdad interesan, de las que luego se estimarán sus efectos.

El experimentador se plantea, entre otras cosas, la posibilidad de lograr mejores estimadores sin tener para ello que aumentar el tamaño del experimento, es decir, sin incrementar los costes. Así surge el *diseño óptimo de experimentos*, que se implica más bien, y sobre todo, en la elección de las unidades experimentales sobre las que realizar el estudio, de manera que los estimadores que se obtengan a partir de éste sean los mejores posibles, es decir, los de mínima varianza. También, y de modo simultáneo, se buscarán covarianzas pequeñas.

Éste va a ser el objetivo del *diseño óptimo*: los mejores diseños serán aquéllos que de alguna manera 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, y así no habrá en general un diseño óptimo “universal” para un experimento, sino respecto de cada criterio. Se ha desarrollado así una amplia gama de criterios de optimización acordes a las distintas perspectivas de la experimentación. Éstos se verán con detalle más adelante en este mismo capítulo.

2.1 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 Kirstine Smith. Propuso un criterio para la regresión polinomial, que más tarde, en 1959, fue llamado G -optimización (*Generalized Variance*) por Kiefer y Wolfowitz. 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 mínimos cuadrados, cuya inversa es proporcional a la 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, y no emplean los llamados criterios “alfabéticos”. Desde este punto de vista la generalización a funciones no polinómicas se hace problemática. Por otro lado J. Kiefer propondrá el empleo de funciones de la matriz de dispersión como posibles criterios de optimización. Desarrolla así la llamada teoría convexa de diseños aproximados. La novedad estriba en considerar un diseño como medida de probabilidad. Algunos de sus seguidores son Atwood, Covey-Crump, Silvey, Fedorov, Karlin, Studden, Whittle, Wynn,...

En 1943 Wald establece el criterio de maximización del determinante de la matriz de información. Más tarde Kiefer y Wolfowitz le darán el nombre de D -optimización (*Determinant*) y extenderán su utilización al modelo de regresión más general. Es éste el más popular de todos los criterios. Diez años después Chernoff 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 (A -optimización, A -

verage), que ya había utilizado Elfving en 1952. El propio Elfving (1952) aborda el problema de la optimización de una combinación lineal de los parámetros introduciendo el criterio de c -optimización y proporcionando incluso un método gráfico para el cálculo del diseño c -óptimo. De la Garza (1954) 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 establece un nuevo criterio de optimización que consiste en maximizar el mínimo autovalor de la matriz de información, y que será llamado E -optimización (*Eigenvalues*). Son estos, entre otros, los llamados criterios alfabéticos, por la denominación que se les ha ido dando.

En 1958 Hoel 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 que establecerán Kiefer y Wolfowitz en 1959. También en 1958 Guest 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, como ya se ha comentado, y el teorema de equivalencia entre los criterios de D y G -optimización. 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 además (1959b) 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. Poco después (1960-61), dan ejemplos de la construcción de diseños D -óptimos, pero no establecen un método general. De modo independiente Fedorov (1972) y Wynn (1970) 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. El libro de Fedorov es publicado en ruso en 1969. En 1970 Wynn publica un artículo con el desarrollo del algoritmo de construcción de diseños D -óptimos. Cuando en 1972 se publica la traducción hecha por Studden aparece allí reflejado básicamente el mismo

algoritmo. Hoy día se admite la producción independiente del algoritmo por los dos autores.

Por esas fechas Box y Lucas (1959b) 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 elipsoide de confianza de las estimaciones de los parámetros.

En 1965 Box y Hunter 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 los algoritmos sugeridos por Fedorov y Wynn a partir del teorema de equivalencia. Karlin y Studden (1966), extienden y simplifican los resultados de optimización parcial, en particular cuando la matriz de información es singular. Draper y Hunter (1967), discutieron el problema de seleccionar distribuciones de parámetros a priori con el objeto de obtener diseños para modelos no lineales. Atkinson y Hunter (1968) extendieron los resultados de Box y Lucas al caso en que el diseño toma más de m puntos. Box (1968-70) da algunos resultados adicionales para modelos no lineales. Por aquel entonces Atwood (1969) obtiene mejores resultados en optimización parcial, y estudia el papel de la simetría en el diseño óptimo.

El cálculo del diseño óptimo no es fácil de implementar. Nalimov y otros (1970), y Box y Draper (1971) 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 podrá aplicar el algoritmo de Fedorov y Wynn. Otros autores desarrollarán algoritmos para su obtención, pero sin garantizar la convergencia. En 1972 Wynn 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 dan una interpretación geométrica del diseño óptimo y plantean un algoritmo para obtener un diseño D -óptimo en el espacio dual. El propio Titterington (1975) ahondará en los aspectos geométricos del D -óptimo. Whittle (1973) generaliza el teorema de equivalencia para cualquier función criterio convexa, y al mismo tiempo White (1973) lo extiende a diseños para modelos no lineales. Kiefer (1974) da resultados de equivalencia para otros criterios. Atkinson y Fedorov introducen en 1975 el criterio T para la discriminación entre modelos. Wu y Wynn

(1978b) dan condiciones generales para la convergencia de los algoritmos para la obtención del diseño óptimo.

Hill (1980) 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. Currie (1982) compara diversos diseños para estimar los parámetros en la ecuación de Michaelis-Menten, frecuentemente utilizada en cinética de enzimas. Abdelbasit y Placket (1983) 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), Pázman (1980),... El artículo de Ash y Hedayat (1978) 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. Los libros de Fedorov (1972), Silvey (1980), Pázman (1986) y los más recientes de Atkinson y Donev (1992) y Pulkenseim (1993) 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 (véase Brown, 1985). Dicha colección es de un inestimable valor para los investigadores en esta materia. A la memoria de Kiefer está dedicado el libro de Shah y Sinha (1989). Entre los últimos en aparecer destacan el de Schwabe (1996), que se centra en modelos multifactoriales, y el de Fedorov y Hackl (1997), donde se introducen temas y modelos de interés en la investigación actual en diseño óptimo. En España es de recientísima aparición el libro de Rodríguez Torreblanca y Ortiz Rodríguez (1999), el que posiblemente sea el primer volumen en español dedicado íntegramente a Diseño Óptimo de Experimentos .

El uso de la informática en los diversos campos de la estadística supone un avance considerable. En particular, en el diseño óptimo de experimentos, entre los primeros que investigan sobre esto están Box y Hunter (1965) 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 para la búsqueda de diseños D -óptimos. En 1980 Galil y Kiefer hacen algunas modificaciones. Welch (1982) desarrolla un nuevo programa más completo. Más tarde Atkinson y Donev (1992) proponen un programa en FORTRAN para diseños exactos. Paralelamente en 1974 Snee y Marquardt desarrollan el programa XVERT para el diseño óptimo en mixturas de modelos. En 1983 Nigam y Gupta proponen una nueva versión de este algoritmo. Hardin y

Sloane crean en 1994 el programa GOSSET, capaz de buscar diseños óptimos respecto de algunos criterios muy utilizados para los modelos polinómicos de grados bajos, con multitud de variables de tipos distintos y restricciones de varias clases. En 1995 Rasch y Darius realizan una revisión de los programas que se pueden utilizar para distintos aspectos del diseño, tanto creados específicamente para este objetivo como formando parte de otros paquetes más generales. El programa SAS incluye en sus últimas versiones un módulo dedicado al cálculo de diseños óptimos.

2.2 Etapas

Como ya se ha dicho, partiremos de un modelo lineal, y lo que nos va a interesar no es tanto estimar los parámetros del modelo, como diseñar un experimento que permita que esa estimación sea óptima.

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. La elección del modelo se abordará con mayor amplitud más adelante, aunque estrictamente hablando no es uno de los objetivos del diseño óptimo, ya que siempre supondremos que dicho modelo ya ha sido escogido.

Una vez determinado el modelo más adecuado al caso que nos ocupe, habrá que reflexionar sobre la manera en que nos va a interesar optimizar los resultados. Como ya hemos dicho, lo ideal sería que los estimadores que se obtuvieran con el diseño tuvieran varianza mínima. Pero en general el modelo tendrá más de un parámetro, y tendremos que plantearnos de qué manera queremos minimizar las varianzas de los estimadores que entran en juego: ¿nos interesa alguno especialmente, o un subconjunto de ellos?, ¿es primordial minimizar la suma de sus varianzas, o quizá la mayor de ellas...? ¿vamos a tener en cuenta la posible correlación entre ellos (covarianzas)? ¿sería interesante tener en cuenta las unidades de medida de las magnitudes a estimar, para evitar en lo posible una descompensación excesiva entre unas y otras?... Según las preferencias u opiniones subjetivas u objetivas de los experimentadores respecto de cada caso, se seleccionará uno u otro criterio de optimización. Éste tema será tratado en los capítulos siguientes.

Finalmente llega el momento de los cálculos: formulado un problema de estimación para el cual tenemos dado el modelo y elegido el criterio de optimización más adecuado, hay que calcular el diseño óptimo. En algunos casos podremos hacerlo exactamente, pero en otros habrá que conformarse con la aplicación de algoritmos numéricos generales o especialmente diseñados para ese problema particular que nos lleven a la obtención de un diseño “suficientemente” bueno. Eso necesitará la medición de la distancia que separa el diseño óptimo de los diseños que vamos obteniendo en cada paso, para tomar una decisión acerca del momento de parada. Se profundizará en este tema más adelante.

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 los intereses reales.
3. Utilización de un algoritmo u otro procedimiento para el cálculo del diseño óptimo.

2.3 Conceptos básicos

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*, θ , 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 es parcialmente desconocida:

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

Un buen diseño tratará de reducir al mínimo la influencia de la inestabilidad de las condiciones. También se podría expresar:

$$y(x) = \eta(x) + \varepsilon_x, x \in X, \quad E(\varepsilon_x) = 0$$

La función η se conoce como *superficie respuesta* o *función de regresión*. Habitualmente supondremos que dicha función es parcialmente conocida, es decir, que está dentro de un conjunto paramétrico de funciones

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

donde los parámetros $\alpha^t = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$ son desconocidos y su especificación determina totalmente a η . Del mismo modo, la varianza podría ser parcialmente conocida, dependiente de estos mismos parámetros u otros, que se introducirían en el modelo para la búsqueda del diseño óptimo.

El caso más interesante que trataremos especialmente es el ya mencionado modelo lineal (en los parámetros) con observaciones incorreladas. Supondremos que X es un espacio métrico y compacto. Como ya hemos dicho esta restricción no es transcendental ya que frecuentemente se presenta en la práctica. El problema se plantea de la forma siguiente:

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

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

Por tanto el *modelo de regresión lineal con observaciones incorreladas* viene determinado por una terna (X, Θ, σ) donde X es un espacio métrico y compacto, Θ es un \mathbb{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 Θ , y redefinir el problema con nuevos parámetros. El espacio vectorial Θ será llamado *espacio de estados*. Por su parte σ será una función positiva y continua. Además las observaciones $y(x), x \in X$, son variables aleatorias incorreladas.

2.3.1 Diseños exacto y aproximado

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 de probabilidad 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, ξ , en X con soporte finito.

Cabría aún una definición más general del diseño como una medida de probabilidad cualquiera, en cuyo caso aparecerían también *diseños continuos* (Atkinson, 1992). Aunque en ocasiones estos diseños podrían ser poco viables en la práctica, son convenientes para demostrar ciertas propiedades. Además eligiendo N suficientemente grande, podremos aproximar un diseño de estas características a uno exacto, tomando un número cercano a $N \xi(x)$ observaciones en el punto x . Por supuesto estas aproximaciones proporcionan diseños tanto mejores cuanto mayor sea N , resultando peligrosas para tamaños pequeños. Con respecto a su eficiencia véase Imhof, López-Fidalgo y Wong (2000), que extienden los diseños exactos conocidos en regresión polinomial y dan cotas de eficiencia para diseños aproximados redondeados con métodos tradicionales.

El diseño concentrado en los puntos x_1, \dots, x_N , con pesos respectivos p_1, \dots, p_N ($0 \leq p_i \leq 1$ para $i = 1, \dots, N$; $\sum_{i=1}^N p_i = 1$) se denotará por

$$\xi = \left\{ \begin{array}{ccc} x_1 & \dots & x_N \\ p_1 & \dots & p_N \end{array} \right\}$$

y el peso de un punto x_i , p_i , se denotará también como $\xi(x_i)$.

El soporte de un diseño ξ será:

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

2.3.2 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:

$$g : \Theta \longrightarrow \mathbb{R}$$

Supondremos que la aplicación g tiene como matriz asociada en la base $\{f_1, \dots, f_m\}$ de Θ el vector $\mathbf{c}^t = (\mathbf{c}_1, \dots, \mathbf{c}_m)$. Es decir, dado un estado:

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

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

Buscamos 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:

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

Dado un diseño de tamaño fijo N , utilizaremos la siguiente notación:

$$Y = (y(x_1), \dots, y(x_N))^t$$

$$A = (a_1, \dots, a_N)^t$$

$$X = \begin{pmatrix} f_1(x_1) & \dots & f_m(x_1) \\ \dots & \dots & \dots \\ f_1(x_N) & \dots & f_m(x_N) \end{pmatrix}$$

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:

$$\text{Var} \left[\sum_{i=1}^N a_i^* y(x_i) \right] = \min \left\{ \text{Var} [A^t Y] : \begin{array}{l} A \in \mathbb{R}^N, \\ E[A^t Y] = g(\theta), \theta \in \Theta \end{array} \right\}$$

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

$$\sum_{i=1}^N a_i^{*2} \sigma_i^2 = \min \left\{ \sum_{i=1}^N a_i^2 \sigma_i^2 : A \in \mathbb{R}^N, \sum_{i=1}^N a_i \theta(x_i) = g(\theta), \theta \in \Theta \right\}$$

Nos interesa obtener una expresión explícita de estos coeficientes.

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

Definición 2.2 *Sea A una matriz cualquiera de orden m . Se definen los conjuntos siguientes:*

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

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

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

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 2.3 Se define la matriz de información asociada a un diseño aproximado ξ 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 La matriz de información de Fisher, supuesta la normalidad de las observaciones, coincidirá con la matriz de información definida anteriormente:

$$M(\xi) = -E_{\xi} \left[\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \log l(y, \theta, \sigma^2) \right] = E_{\xi} \left[\frac{\partial \log l}{\partial \alpha_i} \frac{\partial \log l}{\partial \alpha_j} \right],$$

donde l es la función de verosimilitud de la muestra. Esta definición será así aplicable a modelos no lineales, pero entonces dependerá de los parámetros que se quieran estimar.

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 ξ 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:

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

La ventaja de esta definición es que $M(\xi)$ no depende del tamaño de la muestra, N , sino de la proporción de observaciones en cada punto. Incluso, si hay homocedasticidad, suele considerarse

$$M(\xi) = \frac{1}{\sigma^2 N} M,$$

de modo que no aparezca ningún parámetro en la matriz de información.

Notación 2.1 El conjunto de todos los diseños en el modelo se denotará por Ξ , 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 Ξ . 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:

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

□

Observación 2.2 *En muchas ocasiones suponer, en el Diseño Óptimo de Experimentos, 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. 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.

Proposición 2.2 *Se verifica que*

1. *La matriz de información es simétrica y semidefinido positiva.*
2. *Si ξ tiene menos de m puntos en su soporte, entonces $\det M(\xi) = 0$.*
3. *Se puede deducir la siguiente expresión explícita para el determinante de la matriz de información:*

$$\det M(\xi) = \sum_{k_1 < \dots < k_m} \xi(x_{k_1}) \cdots \xi(x_{k_m}) \det[\{f_i(x_{k_j})\}]^2$$

Demostración:

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

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

Veremos más adelante que la inversa de la matriz de información es proporcional a la matriz de covarianzas, que prueba directamente la condición de definido positiva en el caso no singular.

2. Supongamos que ξ 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.
3. Para simplificar 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:

$$\begin{aligned} m_{ij} &= \sum_{x_k \in X_\xi} f_i(x_k) f_j(x_k) \xi(x_k) \\ a_{ik} &= f_i(x_k) \xi(x_k), b_{jk} = f_j(x_k) \end{aligned}$$

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:

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

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). La expresión del determinante para m puntos era bien conocida. La fórmula más general que se proporciona aquí se debe a Ardanuy y otros (1999). \square

Ejemplo 2.1

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:

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

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, que se debe 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 $Aw_u = u$. Por tanto A^+ verifica $A^+AA^+ = A^+$ y $AA^+A = A$, de modo que a su vez A es la g -inversa de Penrose de A^+ .

Proposición 2.3 *Se cumplen las siguientes propiedades:*

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

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

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

$$\text{Cov}(\hat{g}_1, \hat{g}_2) = N^{-1} \mathbf{c}_1^t M^{-}(\xi) \mathbf{c}_2,$$

Observación 2.4 Estas expresiones no dependen de la g -inversa elegida. En efecto, denotando $A = M(\xi)$, si $c \in \mathcal{M}[M(\xi)]$ existe z tal que $c = Az$, luego

$$c^t A^{-} c = z^t A A^{-} A z = z^t A z,$$

que no depende de A^{-} . Del mismo modo, si $c_1, c_2 \in \mathcal{M}[M(\xi)]$ existirán z_1, z_2 tales que

$$c_1^t A^{-} c_2 = z_1^t A z_2,$$

que de nuevo no depende de la inversa generalizada elegida A^{-} .

Observación 2.5 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 2.4 Dado un diseño ξ y un funcional g , se define la varianza generalizada de g respecto del diseño ξ como:

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

que puede también escribirse de la forma:

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

Y se define la covarianza generalizada de dos funcionales g_1 y g_2 respecto de ξ como:

$$\text{Cov}_{\xi}(g_1, g_2) = \mathbf{c}_1^t M^{-}(\xi) \mathbf{c}_2$$

Estamos interesados en buscar ξ de manera que se hagan pequeñas las varianzas y covarianzas de este tipo. A este fin van encaminadas las siguientes proposiciones, cuya demostración por brevedad se omite (véase Pázman (1986) para más detalles).

Proposición 2.4 Si $M(\xi) \geq M(\eta)$, es decir $M(\xi) - M(\eta)$ es semidefinido positiva, entonces $Var_{\xi}g \leq Var_{\eta}g$ cualquiera que sea el funcional lineal g definido en el espacio de estados. Recíprocamente, si $Var_{\xi}g \leq Var_{\eta}g$ 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 $Var_{\xi}g = Var_{\eta}g$, para cada funcional lineal, g , definido en el espacio de estados.

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

$$M(\xi) \geq M(\eta) \iff Var_{\xi}g \leq Var_{\eta}g, \text{ para cada funcional } g$$

Además:

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

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

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

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

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

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

puede ser expresado como una combinación convexa de a lo sumo $k + 1$ puntos del conjunto T . Es decir, dado un elemento q de $\text{conv}(T)$, 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 plano. Si T es un conjunto de cuatro puntos no alineados en el plano, cualquier punto de $\text{conv}(T)$ se encuentra en un triángulo formado por tres de ellos.

Proposición 2.8 *Si T es un conjunto acotado de \mathbb{R}^k entonces $\text{conv}(T)$ 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 2.8 $\mathcal{M} = \text{conv}(\mathcal{S})$ también lo será. \square

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 y Studden (1966), pg 787.

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

Demostración: Definimos el vector:

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

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

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

□

2.4 Criterios de optimización

2.4.1 Introducción

¿Qué queremos decir con diseño óptimo de un experimento? ¿Cómo entender la expresión “el mejor de los diseños posibles”? Estas preguntas no tienen una respuesta unívoca. 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. Vamos a dar ahora una primera definición de lo que va a ser una función criterio. Más tarde daremos una definición más restrictiva. Para diferenciar una de otra, a la segunda la llamaremos función criterio convexa.

Previamente vamos a definir algunas propiedades de las funciones

$$\phi : NND(m) \longrightarrow \mathbb{R},$$

donde $NND(m)$ denota el conjunto de matrices semidefinido positivas de orden m . En lo que sigue C y D serán dos matrices de $NND(m)$. Para mayor detalle puede consultarse Pukelsheim (1992), cap V, donde se encuentra asimismo la demostración de los Lemas 2.1 y 2.2.

Definición 2.5 *La función ϕ se dice que es isotónica si verifica $\phi(C) \geq \phi(D)$ para $C \geq D$, donde la relación \geq entre matrices se refiere al orden de Loewner.*

Definición 2.6 *Diremos que ϕ es cóncava (convexa) si verifica*

$$\phi(\alpha C + (1 - \alpha)D) \geq_{(\leq)} \alpha\phi(C) + (1 - \alpha)\phi(D)$$

para $0 \leq \alpha \leq 1$, $C, D \geq 0$.

Definición 2.7 La función ϕ será homogénea positiva si $\phi(\delta C) = \delta\phi(C)$ para cualesquiera $\delta > 0$, $C \geq 0$.

Definición 2.8 La función ϕ es superaditiva cuando

$$\phi(C + D) \geq \phi(C) + \phi(D)$$

para cualesquiera $C, D \geq 0$

Además se verifican las siguientes propiedades:

Lema 2.1 Para cualquier función homogénea positiva $\phi : NND(m) \rightarrow \mathbb{R}$, las siguientes condiciones son equivalentes:

1. ϕ es superaditiva.
2. ϕ es cóncava.

Lema 2.2 Para cualquier función $\phi : NND(m) \rightarrow \mathbb{R}$ homogénea positiva y superaditiva, las tres siguientes condiciones son equivalentes:

1. ϕ es no negativa.
2. ϕ es isotónica.
3. ϕ es no negativa, y positiva en $PD(m)$ (matrices definido positivas de orden m), o bien ϕ es idénticamente cero.

2.4.2 Funciones criterio y sus propiedades

Definición 2.9 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 Φ -optimización. Un diseño que minimice $\Phi[M(\xi)]$ se denominará diseño Φ -óptimo.

Proposición 2.10 Toda función criterio Φ , 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 . \square

Observación 2.6 Puede ocurrir que dos funciones criterio den lugar a un mismo criterio de optimización, es decir, que produzcan los mismos diseños óptimos. Es precisamente lo que demostrará el teorema de equivalencia para algunos criterios que veremos más adelante. Desde luego, dado un diseño óptimo, todos los diseños que tengan asociada esa misma matriz de información serán óptimos. Pero también otras matrices de información podrían corresponder a diseños óptimos. Algunos criterios permitirán esto, otros no.

Por el teorema de Caratheodory siempre existirá un diseño óptimo con $1 + m(m + 1)/2$ puntos o menos en su soporte. Para los criterios globales, las matrices óptimas deben ser regulares, de modo que por la Proposición 2.2 un diseño óptimo ha de tener al menos m puntos en su soporte.

Notación 2.2 Llamaremos $\mathcal{M}_+ = \{M \in \mathcal{M} : \det M > 0\}$. $\mathcal{L}(\mathcal{M})$ es el subespacio vectorial generado por el conjunto \mathcal{M} en el espacio vectorial de las matrices simétricas de orden m . Si Φ es una función criterio definimos los conjuntos:

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

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

Proposición 2.11 *Si Φ es una función convexa entonces el conjunto Ξ_{Φ}^* es convexo.*

Demostración: Supongamos que:

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

entonces, por ser Φ 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 2.7 *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:*

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

Aunque la función criterio global sea continua, la función criterio parcial correspondiente no siempre es continua en todo \mathcal{M} . Además el diseño Φ -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 tamaño $s \times s$ de la inversa de la matriz de información.

A partir de los criterios que se estudiarán más adelante pueden definirse criterios parciales de manera natural. Una referencia de ellos puede encontrarse, entre otros libros y artículos en Pázman (1986).

Definición 2.10 *Se dice que Φ es una función criterio estrictamente decreciente cuando las condiciones $M \geq N$ y $M \neq N$ implican $\Phi(M) < \Phi(N)$.*

Proposición 2.12 *Si Φ es una función criterio estrictamente decreciente entonces siempre se puede conseguir que el número de puntos del diseño Φ -óptimo esté comprendido entre m y $m(m+1)/2$.*

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 $m(m+1)/2$ matrices de información de diseños unipuntuales. En efecto, por el teorema de Caratheodory sabemos que si ξ es un diseño tal que $M(\xi)$ está en la frontera de \mathcal{M} entonces existen no más de $m(m+1)/2$ diseños unipuntuales de modo que $M(\xi)$ se puede poner como combinación convexa de ellos. En otras palabras, existe un diseño, η , con soporte en no más de $m(m+1)/2$ puntos tal que $M(\xi) = M(\eta)$. Si demostramos que todo diseño Φ -óptimo, ξ^* , 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 α tal que la matriz de información $(1-\alpha)M(\xi^*) = M(\mu)$ sigue estando en el conjunto \mathcal{M} . Ahora bien $\Phi[M(\mu)] = \Phi[(1-\alpha)^m M(\xi^*)] < \Phi[M(\xi^*)]$ (ya que $(1-\alpha)^m M(\xi^*) < M(\xi^*)$), que es contradictorio con el hecho de que ξ^* es un diseño Φ -óptimo. \square

Definición 2.11 *Se dice que una función criterio Φ es homogénea si verifica $\Phi(\delta M) = \delta^{-1}\Phi(M)$ siendo δ cualquier número real positivo.*

Observación 2.8 *Nótese la diferencia entre la homogeneidad definida para las funciones criterio y la que se acaba de definir. Aunque aparentemente discrepantes, ambas tienen perfecto sentido si se observa que las funciones criterio siempre actúan sobre la inversa de la matriz de información, que es sobre la que se aplican.*

Notación 2.3 *En lo que sigue denotaremos indistintamente $\Phi[M(\xi)]$ ó $\Phi(\xi)$.*

Definición 2.12 *Se define la eficiencia de un diseño ξ respecto de una función criterio homogénea positiva inversa Φ como*

$$eff_{\Phi}(\xi) = \frac{\Phi(\xi^*)}{\Phi(\xi)},$$

donde ξ^ es el diseño óptimo para el criterio Φ .*

Observación 2.9 *En términos prácticos, supongamos que el diseño ξ es ahora un diseño exacto de tamaño N . Hemos visto que $NM(\xi) = M$, de modo que $\Phi[M(\xi)] = N\Phi(M)$. Así, para conseguir una misma eficiencia con ambos diseños tendríamos que tomar N^* observaciones con el diseño ξ^* , de modo que*

$$1 = \frac{\Phi(M^*)}{\Phi(M)} = \frac{N\Phi(\xi^*)}{N^*\Phi(\xi)} \Leftrightarrow \frac{N^*}{N} = \text{eff}_\Phi(\xi).$$

Así por ejemplo, si la eficiencia de un diseño es del 50% entonces bastará tomar la mitad de observaciones con el diseño óptimo para obtener la misma precisión que con el diseño original.

Ahora, para la definición y propiedades de algunas funciones criterio necesitaremos las siguientes definiciones:

Definición 2.13 *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 2.14 *Sea Φ una función definida en un entorno de la matriz A en el espacio $\mathcal{M}_m(\mathbb{R})$. Se define el gradiente de Φ en la matriz A como la matriz de componentes:*

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

Damos ya la definición de las funciones criterio más utilizadas. Aunque aparecen en diversas obras, la mayoría de las definiciones y proposiciones que siguen pueden encontrarse en Pázman (1986).

D-optimización

Definición 2.15 *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 2.13 *Se verifica*

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

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

Observación 2.10 *La gran ventaja de esta función criterio radica en la facilidad de cálculo respecto al resto. Tiene además una sencilla e intuitiva interpretación geométrica, ya que las longitudes de los ejes del elipsoide de confianza de las estimaciones de los parámetros del modelo son proporcionales a las raíces cuadradas de los valores propios de la matriz de covarianzas. En otras palabras, el diseño D-óptimo, al minimizar el determinante de la matriz de información, minimiza el volumen de la región de confianza de las estimaciones de los parámetros (véase por ejemplo Atkinson y Donev (1992), pgs 42 y 48-53). También suele considerarse este criterio en la forma:*

$$\Phi[(M(\xi))] = \det M^{-\frac{1}{m}}(\xi).$$

De este modo se consigue homogeneidad, que permite el cálculo de la eficiencia. El uso del logaritmo permite dar una prueba más sencilla de la convexidad y proporciona un gradiente más simple.

A-optimización

Definición 2.16 *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 2.14 Φ_A *tiene las propiedades siguientes:*

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

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

Observación 2.11 *La interpretación estadística del criterio de A-optimización es clara: el A-óptimo minimiza el promedio de las varianzas de los estimadores de los parámetros, pero, a diferencia del caso anterior, no se fija en las covarianzas entre ellos.*

G-optimización

Definición 2.17 *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:

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

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 2.15 *La función Φ_G es:*

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

Observación 2.12 *Para cada x $\text{Var}_\xi g_x$ es la varianza generalizada de la predicción de la respuesta en x . Por tanto, este criterio busca optimizar la predicción del modelo.*

Criterios minimax

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

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

E-optimización

Definición 2.18 *El criterio de E-optimización viene definido por la función criterio siguiente:*

$$\Phi_E[M(\xi)] = \sup\{\text{Var}_\xi g : \|g\| = 1\}, \xi \in \Xi$$

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

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

En efecto, $\Phi_E[M(\xi)] = \infty$ si, y sólo si existe algún funcional lineal g que no es estimable para ξ , 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:

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

Por otra parte todo vector g tal que $\|g\| = 1$ puede descomponerse de la forma:

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

y por tanto

$$g^t M^{-1}(\xi)g = \sum_{i,j} c_i u_i^t M^{-1}(\xi)u_j c_j \leq \lambda_\xi^{-1},$$

ya que $u_i^t M^{-1}(\xi)u_j = 0$.

Proposición 2.16 Φ_E *tiene las propiedades siguientes:*

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

Observación 2.13 *Como ya se ha mencionado, las raíces de los autovalores son proporcionales a los ejes del elipsoide de confianza de los estimadores de los parámetros. Por tanto, este criterio minimiza el mayor de los ejes. El criterio de D-optimización podría conducir a un elipsoide de volumen mínimo pero con alguno de los ejes excesivamente grande.*

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:

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

Este criterio, ampliamente empleado por Jacroux (1983,87,88) para diseños experimentales de ANOVA ha sido últimamente utilizado en regresión lineal y logística (López Fidalgo (1992), Dette (1993,97), Torsney y López Fidalgo (1995), López Fidalgo, Torsney y Ardanuy (1998)). Ford (1976) dio la derivada direccional.

L-optimización

La siguiente definición, que se debe a Atwood (1976) pg 1125, proporciona una clase de criterios lineales, entre los que se cuenta el de A-optimización.

Definición 2.19 *El criterio de L-optimización viene definido por la función criterio siguiente:*

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

donde W es una matriz definido positiva de orden m .

Proposición 2.17 Φ_L tiene las propiedades siguientes:

1. Φ_L es continua en \mathcal{M} .

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

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

Φ_p -optimización

Definición 2.20 El criterio de Φ_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:

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

L_p -optimización

Este caso viene a generalizar gran parte de los criterios de optimización más usuales, y es una combinación de L - y Φ_p -optimización:

Definición 2.21 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 2.18 *Si p es un entero positivo, entonces Φ_{L_p} tiene las propiedades siguientes:*

1. Φ_{L_p} es continua en \mathcal{M} .
2. Φ_{L_p} es convexa en \mathcal{M} y estrictamente convexa en \mathcal{M}_+ .
3. En las matrices en que Φ_{L_p} es finita, también es diferenciable. Además su gradiente, $\nabla\Phi_{L_p}[M(\xi)]$, es:

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

V-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_0 , este criterio se interesará en minimizar el valor esperado de dichas varianzas sobre el conjunto X . Suponiendo que μ es una medida de probabilidad sobre X se define la función criterio como:

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

c-optimización

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

Definición 2.22 *El criterio de c-optimización para un vector c de dimensión m viene definido por la función criterio siguiente:*

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

Se toma esta definición puesto que la varianza de $c^t\beta$ es proporcional a $c^tM^{-1}(\xi)c$. La desventaja de los diseños c -óptimos es que son singulares con

cierta frecuencia. Elfving (1952) propone un método gráfico para el cálculo del diseño c -óptimo, aplicable sobre todo en el caso biparamétrico:

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

Suponemos que se toma un diseño exacto cualquiera, ξ , concentrado en l puntos x_1, \dots, x_l , 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 l puntos del plano real definidos de la forma $X_1 = (f(x_1), g(x_1)), \dots, X_l = (f(x_l), g(x_l))$. La media de las observaciones en cada punto x_i tiene la siguiente ecuación de regresión:

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

donde $\bar{\varepsilon}_i = \frac{\eta_i}{\sqrt{p_i}}$ tiene varianza σ^2/N_i , y por tanto $\text{Var}(\eta_i) = \sigma^2/N$. Por simplicidad de notación supondremos que $\sigma^2/N = 1$, que no significa ninguna merma en la generalidad de los resultados. De este modo $\text{Var}(\eta_i) = 1$. Nos centraremos en estimadores de la forma $\hat{\varphi} = \sum_{i=1}^l a_i \bar{y}_i$ (\bar{y}_i es el estimador lineal más eficiente de $f(x_i)\alpha + g(x_i)\beta$). Para que sea centrado tiene que cumplir que

$$E[\hat{\varphi}] = \sum_{i=1}^l a_i [f(x_i)\alpha + g(x_i)\beta] = c_1\alpha + c_2\beta,$$

es decir, se ha de satisfacer la ecuación vectorial:

$$\sum_{i=1}^l a_i X_i = \mathbf{c} \tag{2.2}$$

La varianza del estimador es

$$\text{Var}(\hat{\varphi}) = \sigma^2 \sum_i \frac{a_i^2}{N_i} = \frac{\sigma^2}{N} \sum_i \frac{a_i^2}{p_i} = \sum_i \frac{a_i^2}{p_i}.$$

Para minimizarla y que se cumpla a la vez la condición $\sum p_i = 1$ (los p_i no aparecen en la condición (2.2)) utilizaremos los multiplicadores de Lagrange. Así habrá que hallar p_1, \dots, p_l, λ que minimicen

$$\sum_i \frac{a_i^2}{p_i} - \lambda(1 - \sum_i p_i),$$

de lo que resulta

$$p_i = \frac{|a_i|}{\sum_j |a_j|} \implies \text{Var}(\hat{\varphi}) = \left(\sum_{i=1}^l |a_i| \right)^2.$$

Nuestro objetivo ahora es, por tanto, tomar los coeficientes a_i que hagan mínima la varianza anterior bajo la condición (2.2), esto es,

$$\begin{aligned} & \min \left\{ \sum_i |a_i| : \sum_{i=1}^l a_i X_i = c \right\} = \\ & \min \left\{ \sum_i |a_i| : \sum_{i=1}^l w_i [\text{sgn}(a_i) X_i] = t \mathbf{c} \right\}, \end{aligned}$$

donde

$$w_i = \frac{|a_i|}{\sum_j |a_j|}, \quad t = \frac{1}{\sum_j |a_j|}.$$

En consecuencia buscamos el mayor t de modo que $t\mathbf{c}$ esté en el cierre convexo de x_1, \dots, x_l :

$$\mathbf{c}^* = t\mathbf{c} = \sum_i a_i^* X_i = \sum_i t a_i X_i,$$

es una combinación convexa de los vectores $\pm X_1, \dots, \pm X_l$.

La varianza alcanzará su mínimo cuando el extremo del vector \mathbf{c}^* coincida con la intersección del vector \mathbf{c} , o su prolongación, con el polígono dado por el cierre convexo G de los vectores $\pm X_1, \dots, \pm X_l$. La varianza es entonces

$$\text{Var}(\hat{\varphi}) = \frac{1}{t^2} = \frac{\|\mathbf{c}^*\|^2}{\|\mathbf{c}\|^2},$$

y el diseño óptimo

$$\left\{ \begin{array}{ccc} x_1 & \cdots & x_l \\ \frac{|a_1|}{\sum_j |a_j|} & \cdots & \frac{|a_l|}{\sum_j |a_j|} \end{array} \right\}.$$

La Figura 2.1 representa este procedimiento para el modelo biparamétrico. El punto de corte del vector \mathbf{c} , o su prolongación, con la frontera de G estará situado en el segmento que une dos de los vectores $\pm X_1, \dots, \pm X_l$ (en este caso son X_2 y $-X_1$). Así $\mathbf{c}^* = a_1^*(-X_1) + a_2^*X_2$ (con $a_1^* + a_2^* = 1$). Como, por otra parte, \mathbf{c}^* se podría expresar en la forma $X_2 + \lambda(-X_1 - X_2) =$

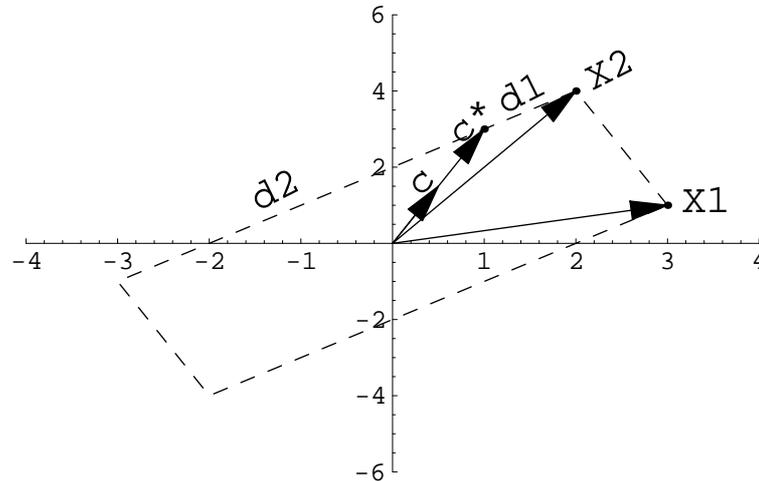


Figura 2.1: Método gráfico para el cálculo del c -óptimo

$\lambda(-X_1) + (1-\lambda)X_2$, $0 \leq \lambda \leq 1$ (por el hecho de estar en el segmento que une los dos vectores), resulta inmediatamente que $a_1^*(= \lambda)$ es proporcional a la distancia d_1 de c^* a X_2 sobre el segmento, y a_2^* proporcional a la distancia d_2 de c^* a X_1 sobre ese segmento, que es el método propuesto por Elfving. Esto es aplicable para cualesquiera puntos posibles del espacio del diseño. Por tanto, en la práctica habrá que calcular el cierre convexo de $f(\mathcal{X}) \cup -f(\mathcal{X})$ (\mathcal{X} =espacio del diseño) y la intersección de la recta definida por el vector c con la frontera de dicho cierre.

Observación 2.14 *Este criterio proporciona en particular los mejores diseños para estimar cada uno de los parámetros. Esto es una buena referencia para estudiar la bondad (eficiencia) de un diseño cualquiera para estimar cada parámetro en particular.*

Criterios estandarizados

Detle (1997) propone una manera novedosa de tratar las funciones criterio conocidas. Tiene en cuenta las diferentes magnitudes de los parámetros objeto de estimación, que provocaban una muy distinta influencia de los mismos

en la elección de los mejores diseños. “Retoca” las funciones criterio para evitar esta diferencia, estandarizando las varianzas de los estimadores. Así todos los parámetros quedarán estimados con la misma precisión, independientemente de las unidades de medida en que vengan dados. Explícitamente, si ψ es una función no negativa, homogénea positiva, superiormente semicontinua y cóncava sobre el espacio

$$\mathbb{R}_{\geq}^l = \{a = (a_1, \dots, a_l)^T \in \mathbb{R}^l \mid a_j \geq 0, \quad j = 1, \dots, l\}$$

se dice que ξ es S_ψ -óptimo (ψ -óptimo estandarizado) para los parámetros $\vartheta_{i_1}, \dots, \vartheta_{i_l}$ en el modelo

$$y = x^T \theta + \epsilon$$

si maximiza la función

$$\psi(\xi) = \psi \left[\left\{ \frac{e_{i_1}^T M^-(\xi) e_{i_1}}{e_{i_1}^T M^-(\xi_{i_1}^*) e_{i_1}} \right\}^{-1}, \dots, \left\{ \frac{e_{i_l}^T M^-(\xi) e_{i_l}}{e_{i_l}^T M^-(\xi_{i_l}^*) e_{i_l}} \right\}^{-1} \right]$$

donde e_1, \dots, e_m son los vectores de la base canónica y ξ_j^* es el mejor diseño para la estimación de ϑ_j , esto es, el que minimiza la varianza de su estimador. Dicho de otra manera, ξ_j^* es el c -óptimo para $\mathbf{c} = e_j$, y $e_j^T M^-(\xi) e_j = \Phi_{e_j}[M(\xi)]$, luego se puede poner

$$\psi(\xi) = \psi \left[\text{eff}_{\theta_{i_1}}(\xi), \dots, \text{eff}_{\theta_{i_l}}(\xi) \right]$$

El siguiente teorema muestra la relación entre algunos criterios:

Teorema 2.2 *Se verifican las siguientes relaciones entre las funciones criterio:*

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 Φ_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 V -optimización.
7. Si en el criterio minimax se utiliza la norma $|c|_1 = \sum |c_i|$ entonces obtenemos el criterio MV , utilizando la norma $|c|_2 = (\sum c_i^2)^{1/2}$ se obtiene el criterio E y utilizando la siguiente norma se obtiene el criterio G :

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

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 2, 4 y 5 bastará hacer la sustitución indicada. Para probar el apartado 1 denotemos por $\lambda_1, \dots, \lambda_m$ los autovalores de $M(\xi)$ y sea λ_{i_0} el mínimo de los autovalores. Entonces:

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

Para demostrar el 3 hacemos $H = I$ de modo que tomando logaritmos:

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

Veamos ahora que se cumple el resultado 6. En efecto, tomando en L -optimización $W = \int_X f^t(x)f(x)d\mu$ se tendrá:

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

Para probar el 7 vemos que:

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

siendo e_1, \dots, e_m la base canónica. La segunda igualdad se debe 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 $\{c \in \mathbb{R}^m : |c|_1 = 1\}$, que vienen dados por los vectores de la base canónica. Por otra parte,

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

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

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

□

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

2.5 Teorema de equivalencia

Kiefer y Wolfowitz (1960) demostraron 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 2.3 *Suponiendo que $\sigma(x)$ es constante para todo x de X , un diseño ξ^* 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 2.15 *Tener en cuenta que*

1. Si $\sigma(x)$ no es constante, el teorema de equivalencia se cumple también haciendo una pequeña precisión. En este caso puede demostrarse que ξ^* 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. Atkinson y Donev (1992), pgs 98 y ss, proporcionan un ejemplo. En el modelo cuadrático, tomando como intervalo de diseño el $[-1, 1]$, 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 2.16 Veamos ahora con detalle la interpretación geométrica del criterio de D -optimización. Sobre el elipsoide m -dimensional:

1. El volumen del elipsoide m -dimensional:

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

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 $w = \frac{1}{c} U^{-1} z$ 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 \mathcal{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 -dimensional de radio unidad. Pero

$$(\det U)^2 \det M(\xi) = 1,$$

con lo que queda probado el resultado.

2. Bajo la hipótesis de normalidad de las observaciones $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_m)$ se distribuye de acuerdo a una distribución normal multivariante de modo que la media tiene como componentes los verdaderos valores de los parámetros α y la matriz de covarianzas es proporcional a $M^{-1}(\xi)$. Supongamos sin pérdida de generalidad que $M^{-1}(\xi)$ es la matriz de covarianzas. 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 $v = U^{-1}(\hat{\alpha} - \alpha)$ seguirá una distribución normal de media $\mathbf{0}$ y vector de covarianzas la identidad. De este modo $v^t v$ seguirá una distribución χ_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 α 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 ξ 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 ξ es un diseño D -óptimo se cumple:

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

Además como se vió más arriba el volumen de \mathcal{E}_ξ 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}_ξ conteniendo el conjunto $f(X)$, que es por tanto equivalente a hacer mínimo el volumen del elipsoide de confianza \mathcal{E} . Uno de los inconvenientes de este criterio es que a veces el volumen del elipsoide puede ser minimizado haciendose 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 (1973) proporciona un método para su utilización.

Whittle (1973) y Kiefer (1974) amplían el teorema de equivalencia para funciones criterio más generales. Con el objeto de enunciar estos resultados damos una definición más restrictiva de función criterio. Para ello nos fijaremos en las propiedades que verificaban las funciones criterio vistas anteriormente. Observamos que la mayoría de ellas son convexas y continuas. Con el objeto de salvaguardar estas propiedades daremos la siguiente definición:

Definición 2.23 Diremos que una función real Φ 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_Φ abierto de $\mathcal{L}(\mathcal{M})$ tal que $U_\Phi \supset \mathcal{M}_+$ y Φ está definida, es finita y convexa en U_Φ .

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:

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

lo que significaría una especie de continuidad al pasar de \mathcal{M}_+ a $\mathcal{M}\mathcal{M}_+$.

Observación 2.17 No es necesario que U_Φ sea convexo. Entenderemos entonces que Φ debe ser convexa en los subconjuntos convexos de U_Φ .

Proposición 2.19 *Si U es un abierto de $\mathcal{L}(\mathcal{M})$ y Φ es convexa y finita en U , entonces Φ es continua en U .*

Una demostración de este resultado puede encontrarse en Holmes (1975).

Proposición 2.20 *Cualquier función criterio convexa, con la definición dada, es continua en \mathcal{M} .*

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

\mathcal{M}_+ . □

Puesto que muchas de estas funciones no serán diferenciables conviene dar la definición de derivada direccional de Frechét:

Definición 2.24 *Dada una función real Φ , 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 Φ 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 2.18 *Esta derivada direccional existe siempre gracias a la convexidad de Φ . En efecto, demostraremos que la función:*

$$\varphi : (0, 1) \longrightarrow \mathbb{R} \quad | \quad \beta \longrightarrow \frac{\Phi[(1 - \beta)x + \beta v] - \Phi(x)}{\beta}$$

es no decreciente 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 φ 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 Φ 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 2.4 (teorema general de equivalencia) *Sea Φ una función criterio convexa y ξ^* un diseño tal que:*

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

Entonces son equivalentes:

1. ξ^* es Φ -óptimo.
2. ξ^* es Φ -óptimo local, es decir, para cada diseño ξ la función:

$$[0, 1] \longrightarrow \mathbb{R} \quad | \quad \beta \longrightarrow \Phi[(1 - \beta)M(\xi^*) + \beta M(\xi)]$$

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 Φ -óptimo, tanto si la función criterio es diferenciable como si no lo es. Cuando la función Φ sea diferenciable, se puede dar el siguiente teorema:

Teorema 2.5 *Si Φ es diferenciable en un entorno de $M(\xi^*)$, entonces son equivalentes:*

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

La demostración de estos dos últimos teoremas puede encontrarse por ejemplo en Pázman (1986), pgs 116-117. Disponemos así de un par de condiciones para la búsqueda del soporte del óptimo. Si llamamos

$$\psi(x, \xi) = f^t(x) \nabla \Phi[M(\xi)]f(x) - \text{Tr} M(\xi) \nabla \Phi[M(\xi)],$$

tenemos que $\psi(x, \xi^*)$ se anula en todos los puntos del soporte del óptimo, y por otra parte, la segunda condición del teorema anterior nos dice que $\psi(x, \xi^*) \geq 0$. De aquí concluimos que los puntos del soporte del Φ -óptimo son mínimos de la función $\psi(x, \xi^*)$, y por tanto se verifica

$$\frac{\partial \psi(x, \xi^*)}{\partial x} = 0 \quad \forall x \in X_{\xi^*}.$$

De este modo, cuando se tenga alguna sospecha de cómo puede ser el óptimo, estas dos condiciones podrían ayudar a encontrarlo. Posteriormente habría que comprobar con el teorema de equivalencia que efectivamente es el óptimo.

Observación 2.19 *La función $\psi(x, \xi)$, como ya se ha comentado, resulta muy útil para buscar los puntos para los cuales el tomar una observación contribuye más a aproximarnos al diseño óptimo. Es decir, es una función especialmente “sensible” al soporte del óptimo. Por este motivo una ligera modificación suya es bautizada como función de sensibilidad por algunos autores (véase por ejemplo Fedorov y Hackl (1997), pg 33 y ss, para más detalles en este sentido).*

2.6 Una crítica a la teoría del diseño óptimo de experimentos

Jamás existirá una disciplina que sea del agrado de todo el mundo, y el diseño óptimo no podía ser menos. La Teoría del Diseño Óptimo de Experimentos requiere un formalismo y un grado de matematización alto. Esto hace que cualquier desventaja o inconveniente que pueda tener la aplicación práctica de esta teoría se vea amplificada. Intentaremos mostrar una motivación justificada de la aplicación de esta teoría. Esto servirá además para proporcionar unas pautas y directrices de actuación práctica. Consideraremos en primer lugar algunos de los inconvenientes que se le atribuyen a esta teoría y a su aplicación práctica.

1. La teoría del diseño óptimo de experimentos requiere la elección previa del modelo sin contar aún con las observaciones. Por tanto el diseño óptimo dependerá fuertemente del modelo elegido. Puede ocurrir entonces que el diseño no permita detectar la existencia de un modelo mejor que el que se supuso inicialmente. Por ejemplo, podría ser incapaz de descubrir cierta curvatura en un modelo que inicialmente se supuso de regresión lineal. Incluso, en el caso de que nos permitiera detectar un modelo mejor, el diseño podría ser deficiente para la estimación de los parámetros del mismo.
2. En el caso de modelos no lineales habría que añadir además el problema que lleva consigo la elección de unos valores iniciales, o una distribución inicial, de los parámetros del modelo.
3. Un diseño se dice óptimo siempre de acuerdo a un criterio, que depende fundamentalmente de los objetivos del estudio estadístico. De nuevo el criterio ha de fijarse antes de observar los datos. Así un diseño puede ser óptimo para un criterio de interés, pero muy malo desde otro punto de vista, que puede tener también su importancia. Y lo cierto es que además se han propuesto una gran cantidad de criterios y podrían proponerse otros muchos.
4. Se ha mencionado que la teoría del diseño óptimo ha discurrido en dos corrientes, casi paralelas e irreconciliables. Por una parte, la consideración de diseños exactos, que requiere una maquinaria matemática más compleja, conduce a diseños que, en principio, son realizables en la práctica. Por otro lado los diseños aproximados son más fáciles de obtener al ser posible el uso del teorema de equivalencia. No obstante estos diseños se han de “aproximar” en la práctica por diseños realizables, que podrían resultar muy poco eficientes para tamaños de muestra bajos (véase Imhof, López-Fidalgo y Wong, 2000). A todo esto se une el hecho de que en la práctica no es sencillo obtener diseños óptimos y hay que recurrir frecuentemente a métodos numéricos.
5. Con frecuencia los diseños que se obtienen exigen tomar observaciones en condiciones muy extremas. En muchas ocasiones estas condiciones son inviables por razones prácticas e incluso éticas. Aquí podría aplicarse el famoso dicho de que “lo mejor es enemigo de lo bueno”.

6. Algunos criterios no son invariantes por cambios de escala. Eso significa que se puede estar dando una importancia desmedida a algunos parámetros con respecto a otros. De este modo el diseño óptimo resultante podría ser muy bueno para un parámetro, pero muy malo para otro.

Si se analizan estas críticas con detenimiento podría llegarse fácilmente a la conclusión de que no son inconvenientes exclusivos de la teoría del diseño óptimo de experimentos, sino de los métodos estadísticos en general y de la modelización matemática de la realidad. Todo proceso se debilita en las etapas en que hay que elegir. En cualquier análisis estadístico han de hacerse algunas hipótesis, no siempre fáciles de conseguir o comprobar. La elección del modelo y sus condiciones básicas son un ejemplo de ello. Box (1979) afirma que “los modelos nunca son verdad, pero por fortuna solamente basta que sean útiles”.

1. Si bien es cierto que al problema habitual de especificación del modelo en nuestro caso ha de añadirse que no contamos aún con los datos, también es cierto que en la mayoría de los casos, existe una experiencia, datos retrospectivos, “intuiciones” del investigador, ... que hacen razonable una determinada elección del modelo. No obstante, en la práctica es frecuente que la elección del modelo haya de hacerse de entre un número reducido de modelos. Esto conduce al problema de discriminación de modelos. Por supuesto, desde esta perspectiva también pueden considerarse criterios de optimización adecuados (véase por ejemplo Atkinson y Fedorov, 1975).
2. En el uso de modelos no lineales caben también estudios de robustez y sensibilidad de los valores iniciales elegidos. La teoría bayesiana puede ser de gran utilidad en este terreno y de hecho son ya muchos los trabajos realizados en este sentido (véase una recopilación de los mismos en Chaloner y Verdinelli, 1995).
3. La posibilidad de utilizar criterios combinados o compuestos conduce a diseños razonablemente buenos desde diversos puntos de vista (Cook y Wong, 1994). Teniendo en cuenta que el diseño óptimo desde la perspectiva de un criterio determinado no es único, cabe siempre la posibilidad de elegir aquél que sea también satisfactorio, o incluso óptimo,

para uno o más criterios secundarios. Desde luego la variedad de criterios existentes o en estudio no sólo no es una desventaja, sino que ofrecen un amplio abanico de opciones para la elección del más conveniente a los intereses del investigador.

4. El cálculo de diseños exactos podría reducirse a tamaños de muestra bajos, donde tiene mayor utilidad y su cálculo es todavía asequible. Para tamaños altos los diseños aproximados, para cuyo cálculo se dispone de buenos algoritmos implementables en ordenador, conducen a diseños realizables muy eficientes. Con el uso del ordenador los métodos numéricos se hacen tratables y proporcionan soluciones muy próximas a las exactas.
5. Basta observar muchos de los diseños óptimos que se han obtenido para modelos de uso frecuente para darse cuenta de que efectivamente exigen tomar observaciones en condiciones extremas. Podríamos decir en primer lugar, que si dichos diseños no son realizables en la práctica, al menos son una referencia para medir la eficiencia de los diseños que sí se utilizan habitualmente, o que son también satisfactorios para otros criterios o modelos. Siempre cabe además la posibilidad de calcular diseños restringidos o de definir los límites del dominio experimental de modo que las condiciones de frontera no sean tan extremas y de hecho sean factibles en la práctica (Ardanuy y López Fidalgo (1992) o Dette y Sham (1998)).
6. El problema de la invarianza con respecto a cambios de escala ha sido abordado de una manera satisfactoria por Dette (1997). Mediante los criterios estandarizados, que consideran la eficiencia de un diseño para estimar un parámetro, más que el error de estimación, este problema queda corregido proporcionando diseños igualmente eficientes para todos los parámetros del modelo.

Como se ha visto en la introducción histórica, el diseño óptimo es un campo relativamente nuevo, que está avanzando muy deprisa en estos últimos años. Y, ¿qué parte de la ciencia que se precie no está basada en una teoría fuerte sobre la que construir su edificio? Además, múltiples aplicaciones prácticas exigen la construcción de un buen diseño, y son precisamente ellas las que han impulsado el desarrollo del diseño óptimo y su importancia creciente.

Capítulo 3

Algoritmos

3.1 Algunos algoritmos conocidos

A la hora de enfrentarnos al cálculo del diseño óptimo mediante la realización de un procedimiento iterativo, 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 la etapa previa. 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 3.1 *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:*

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

3.1.1 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?. Para afrontar esta cuestión examinaremos los siguientes resultados, que pueden consultarse por ejemplo en Pázman (1986), pg 118.

Proposición 3.2 *Sea Φ una función criterio convexa. Si μ es un diseño tal que $\Phi[M(\mu)] < \infty$ y para un cierto $\delta > 0$ se tiene:*

$$\partial\Phi[M(\mu), M] \geq -\delta, M \in \mathcal{M} \quad (3.1)$$

entonces:

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

Es decir, si ξ^* es Φ -óptimo entonces:

$$\Phi[M(\mu)] - \Phi[M(\xi^*)] \leq - \inf_{M \in \mathcal{M}} \partial\Phi[M(\mu), M] \quad (3.2)$$

Corolario 3.1 *La desigualdad (3.2) proporciona una cota inferior para la eficiencia del diseño:*

$$eff_{\Phi}(\mu) \geq 1 + \frac{\inf_{M \in \mathcal{M}} \partial\Phi[M(\mu), M]}{\Phi[M(\mu)]}$$

Si además existe el gradiente $\nabla\Phi[M(\mu)]$ entonces se tiene el siguiente resultado a partir del Teorema 2.5:

Proposición 3.3 *Si $\Phi[M(\mu)] < \infty$, $\delta > 0$ y existe $\nabla\Phi[M(\mu)]$ verificándose*

$$f^t(x) \nabla \Phi[M(\mu)] f(x) \geq tr \nabla \Phi[M(\mu)] M(\mu) - \delta \quad \forall x \in X$$

entonces

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

Es decir

$$eff_{\Phi}(\mu) \geq 1 - \frac{tr \nabla \Phi[M(\mu)] M(\mu) - \min_{x \in X} f^t(x) \nabla \Phi[M(\mu)] f(x)}{\Phi[M(\mu)]}$$

Esta acotación es válida para cualquier criterio. En 1969 Atwood había dado una acotación del error cometido al aceptar como D -óptimo un diseño dado:

Proposición 3.4 *Si $\det M(\mu) \neq 0$, entonces:*

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

o tomando logaritmos:

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

3.1.2 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 función

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

en el caso general (sin suponer $\sigma^2(x) = 1$).

Para detener el algoritmo puede utilizarse la Proposición 3.4, que puede expresarse de la forma:

$$eff_D(\xi) \geq \frac{m}{\max_{x \in X} d(x, \xi)} \quad (3.3)$$

siendo ξ un diseño tal que $\det M(\xi) \neq 0$. Damos a continuación algunos de los algoritmos más utilizados:

Algoritmo de paso armónico

Puede encontrarse en Pázman (1986), §V.2.

1. Se elige un diseño inicial ξ_0 tal que, $\det M(\xi_0) \neq 0$.

2. Dado un ξ_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:

$$X \longrightarrow \mathbb{R} \mid x \longrightarrow d(x, \xi_n)$$

3. El proceso se detiene de acuerdo con (3.3).

Algoritmo ascendente

Ahora el segundo paso se mejorará en cada etapa del algoritmo.

1. Se elige un diseño inicial ξ_0 tal que $\det M(\xi_0) \neq 0$.
 2. Dado un ξ_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:

$$X \longrightarrow \mathbb{R} \mid x \longrightarrow d(x, \xi_n)$$

y β_n es el punto donde se alcanza el máximo de:

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

3. El proceso se detiene de acuerdo con (3.3).

Proposición 3.5 *En el proceso anterior se puede calcular β_n a partir de la fórmula:*

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

La demostración de este resultado puede encontrarse en Fedorov (1972), §2.5.

La Proposición 3.1 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.

Algoritmo acelerado

Con este proceso el algoritmo convergerá más rápido a partir de una mejor elección de β_n en cada paso. Puede consultarse en Pázman (1986), pg 144.

1. Se elige un diseño inicial ξ_0 tal que, $\det M(\xi_0) \neq 0$.
2. Dado un ξ_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:

$$X \longrightarrow \mathbb{R} \quad | \quad x \longrightarrow d(x, \xi_n)$$

y β_n es el punto donde se alcanza el máximo de:

$$\left[\frac{\xi_n(x_n)}{1 - \xi_n(x_n)}, 1 \right] \longrightarrow \mathbb{R} \quad | \quad \beta \longrightarrow \det M[(1 - \beta)\xi_n + \beta\xi_{x_n}]$$

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 (3.3).

Proposición 3.6 *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 $\beta_n(1 - \beta_n) = \gamma_n$

Observación 3.1 *En todos estos algoritmos se habla de incorporar en el paso $n + 1$ el punto x_n en el que se alcanza el máximo de la función*

$$\begin{aligned} X &\longrightarrow \mathbb{R} \\ x &\longrightarrow d(x, \xi_n) \end{aligned} .$$

En el caso en que este máximo se alcanzara en varios puntos habría varias posibilidades: o bien se incorporan todos esos puntos, dividiendo entre todos ellos el peso correspondiente, o bien se añade solamente uno de ellos. En este último caso, en los siguientes pasos del algoritmo la función alcanzaría su máximo en los puntos restantes siempre que sean relevantes, con lo que se producirá la incorporación de los mismos igualmente, sólo que en un número mayor de etapas.

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. Para este caso existe el siguiente algoritmo:

1. Se elige un diseño inicial ξ_0 tal que, $\xi_0(x_i) > 0, i = 1, 2, \dots, k$.
2. Dado un ξ_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:

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

3.2 Diseños sin replicación para conjuntos finitos

En ocasiones puede que tengamos que enfrentarnos al hecho de que los puntos observables no pueden ser examinados más de una vez, como por ejemplo cuando se trata de pacientes médicos a los que se somete a algún tratamiento un poco especial, o cuando se trata de pruebas largas, donde la repetición sobre la misma unidad experimental se hace inviable, como en agricultura. Para estos casos habrá que realizar una clase especial de diseños: los *diseños sin replicación*.

Un *diseño sin replicación* es por tanto aquél tal que los puntos que constituyen su soporte pueden aparecer contados solamente una vez. Vamos a examinar estos diseños para el caso de conjuntos finitos. Supondremos la población

$$B_N = \{x_1, \dots, x_N\}$$

estando siempre los observables $x_i \in \mathbb{R}$ ordenados, $x_1 < \dots < x_N$, y llamaremos n -diseño a un subconjunto de la forma

$$X_n = \{x_{i_1}, \dots, x_{i_n}\},$$

donde se supone que n es un número fijo, y que buscamos diseños de tamaño n . Se trata pues de diseños exactos en los que no se admite réplica.

Los resultados que figuran a continuación están basados en Rasch y Hendrix (1997):

3.2.1 Cálculo del *diseño sin replicación óptimo*

El método tradicional es el recorrido de todos los n -diseños posibles, probando la función criterio en todos ellos. Puesto que hay

$$\binom{N}{n}$$

diseños de este tipo, el proceso puede resultar largo y costoso.

Una mejora con vistas a su implementación en ordenador sería utilizar una función recursiva para recorrer igualmente todos los n -diseños, pero siguiendo

un patrón determinado. La ganancia se produce porque al recorrer los diseños de esta forma, vamos hallando la información contenida en cada uno de ellos a partir de la que encontramos en el anterior, en vez de calcular esta información en cada paso. Con esto se consigue un ahorro de cálculos. Las dos maneras de emplear este método son:

- Partir de cero e ir añadiendo puntos hasta formar los n -diseños (*adding*)
- Empezar cogiendo todos los puntos e ir eliminando hasta quedarnos con n (*dropping*). Es mejor que el primer procedimiento cuando $n > N/2$

De cualquiera de estas maneras vamos obteniendo un “árbol” de diseños: a partir de un diseño en cada paso surgen una serie de nuevos diseños (ramas) según el punto que se añada (procedimiento *adding*) o se elimine (caso del método *dropping*). Es claro por otra parte que cuantos más puntos tengamos la oportunidad de disponer para incluir o no en el diseño, tanto mejor podrá ser éste, y por ello se puede introducir la siguiente mejora:

Procedimiento de Ramas y Cotas (Branch and Bound): Una vez obtenido un n -diseño, el valor de la función criterio sobre él lo tomamos como cota. Cuando usamos el método *dropping*, al ir disminuyendo el conjunto de puntos candidatos a formar parte del diseño, aumentará el valor de la función criterio. Por tanto, si al principio de cada “rama” del algoritmo (cuando tenemos más de n puntos en el diseño) calculamos la función criterio y es mayor que la cota, no haría falta buscar en esa rama, pues al eliminar puntos el valor de la función criterio irá aumentando para los diseños formados a partir de esa rama. De esta manera se reduce enormemente la búsqueda. La eficacia de este método dependerá mucho de la ordenación de los puntos, ya que si uno de los primeros n -diseños que encontremos es cercano al óptimo, nos evitaremos buscar en muchas de las siguientes ramas, ya que las cotas serán bajas desde el principio. Este proceso de ordenación no sólo debería hacerse en la raíz del árbol, sino en todas las ramas.

3.2.2 Cálculo de un “diseño sin replicación” razonablemente bueno

Si no es imprescindible encontrar el óptimo, sino solamente un diseño que consideremos lo bastante bueno, se pueden utilizar varios métodos, reduciendo los cálculos:

Métodos de extensión o reducción

Extensión: Hallamos el p -óptimo (siendo $p < n$) por alguno de los métodos anteriores (tradicional, *adding*, *dropping*,...), y luego en cada paso añadimos el punto que reduzca más el valor de la función criterio sobre el diseño de partida.

Reducción: Partimos del diseño $X_N = B_N$ (con soporte en todos los puntos), y en cada paso eliminamos el punto que produzca el menor aumento posible de la función criterio sobre el nuevo diseño X_{N-1} .

Algoritmo Combinado: Consiste en la combinación de estos dos métodos para producir mejoras en la utilización de la memoria del ordenador o en el tiempo de cálculo.

Método del subconjunto

La idea es reducir el tamaño del conjunto de búsqueda B_N para disminuir los cálculos. Nos quedaremos, pues, con un conjunto $B_{N'}$ ($N' < N$) constituido por los puntos que cumplan una cierta condición: a partir de un p -diseño óptimo X_p (entre todas las combinaciones posibles de p puntos de B_N) nos quedaremos con los puntos de B_N que estén “cerca” de los puntos del soporte de X_p . Esta cercanía la fija el experimentador con unos parámetros de entorno que serán elegidos según la cantidad N' de puntos con los que queremos quedarnos.

Métodos de intercambio

Partimos de un n -diseño cualquiera, X_n , y en cada paso se sustituye uno de sus puntos por otro de los $N - n$ no contenidos en el soporte, comprobando que la sustitución efectuada es la que más reduce la función criterio de todas las n ($N - n$) posibles. El resultado obtenido al final es un óptimo local, pero no global, ya que dependerá del diseño de partida. Otras variantes del método consisten en intercambiar varios puntos a la vez; en estos casos, tanto los puntos sustituidos como los incorporados suelen tomarse consecutivos.

3.3 Algoritmo Sistemático

En esta sección introduciremos varios algoritmos más para la obtención del diseño óptimo sobre conjuntos finitos de puntos. Son originales del autor de este trabajo, y están relacionados con el llamado *Algoritmo Local*, debido a López Fidalgo (1991), que se presenta a continuación:

3.3.1 Algoritmo local

La idea de partida consiste en la elección de una serie de diseños iniciales, y la búsqueda del óptimo como una combinación de éstos. De esta manera se lograría reducir bastante el espacio de búsqueda. Esto es:

i) Inicialmente el experimentador tiene r diseños, ξ_1, \dots, ξ_r .

ii) Entonces tenemos que resolver el siguiente problema de optimización:

Encontrar

$$\min_{\alpha_1, \dots, \alpha_r} \phi(\alpha_1 \xi_1 + \dots + \alpha_r \xi_r)$$

sujeto a las restricciones

$$\sum_{i=1}^r \alpha_i = 1 \quad \alpha_1 \xi_1 + \dots + \alpha_r \xi_r \geq 0$$

Si k es el número de puntos de la unión de todos los soportes de estos diseños iniciales, tenemos entonces un problema de programación matemática con k restricciones lineales de desigualdad y una de igualdad. No es posible precisar exactamente el número de diseños que hemos de emplear, pero parece claro que cuantos más diseños usemos conseguiremos una mayor aproximación al óptimo. El número total de puntos de la unión de los soportes es también un factor importante. El diseño obtenido a través de este proceso va a tener su soporte contenido en esta unión.

Por tanto, la situación es la siguiente: \mathcal{M} es un subconjunto del espacio lineal de las matrices simétricas de orden m . Denotemos por $\mathcal{L}(\mathcal{M})$ el subespacio generado por \mathcal{M} en ese espacio, que será por tanto de dimensión menor o igual que $m(m+1)/2$. Si pudiéramos encontrar entre las matrices de \mathcal{M} una base de $\mathcal{L}(\mathcal{M})$, y elegir los diseños iniciales correspondientes estaríamos en una situación óptima para aplicar este método. Además podría darse el caso de que $\mathcal{L}(\mathcal{M})$ fuera el espacio vectorial de todas las matrices simétricas lo cual simplificaría la elección de estos diseños.

Sin embargo el método que proponemos no se centra en el conjunto \mathcal{M} sino en Θ , el conjunto de todos los diseños sobre X , que es bastante más complejo que \mathcal{M} . Así, a pesar de que la matriz de información del óptimo, $M(\xi^*)$ sea combinación de las matrices del sistema generador $M(\xi_1), \dots, M(\xi_r)$,

$$M(\xi^*) = \sum_{i=1}^r \alpha_i M(\xi_i),$$

el diseño óptimo ξ^* no tiene por qué ser combinación lineal de los diseños correspondientes a estas matrices. Aún más: $\alpha_1 \xi_1 + \dots + \alpha_r \xi_r$ ni siquiera tendría por qué ser un diseño.

Cuando imponemos la condición

$$\alpha_1 \xi_1 + \dots + \alpha_r \xi_r \geq 0$$

estamos restringiéndonos más de lo necesario; y esto implica que ya no estamos trabajando con todas las matrices de información. Sin embargo no podemos prescindir de esta condición ya que nuestro propósito es calcular explícitamente el diseño óptimo, y no solamente su matriz de información (problema aún más complejo).

Este procedimiento es válido en aquellos casos en los que sabemos que el óptimo es una combinación lineal de algunos diseños dados. Tiene la ventaja de la rapidez de cálculo, ya que el proceso se termina en una sola etapa; pero debemos tener presente que este método busca el ϕ -óptimo entre las combinaciones lineales de diseños dados “a priori”, y no de todos los posibles, y por tanto es necesario saber con alguna aproximación dónde se encuentra el óptimo. Sería aconsejable coger como diseños iniciales aquellos que experimentalmente hayan dado mejor resultado. Incluso en ocasiones nos interesan diseños de un determinado tipo, que de este modo restringimos convenientemente.

Como se puede observar, no se exige que los coeficientes de los diseños sean positivos. La existencia de solución está garantizada, ya que el ϕ -óptimo siempre existe por ser ϕ continua en el compacto \mathcal{M} . Y también existirá un óptimo si nos restringimos, como en este caso, a un conjunto cerrado de \mathcal{M} .

3.3.2 Algoritmo Sistemático

Generalización del “conjunto base”

Como ya hemos mencionado, el argumento anterior es adecuado si se tiene alguna idea de la situación del óptimo. Cuando no es así, el principal problema reside, por supuesto, en la elección de este “conjunto base” de diseños iniciales. Así pues, para el caso general en el que se carece de datos sobre la posible ubicación del óptimo, se reduce la selección a los diseños unipuntuales, diseños “antecedentes” de todos los demás, y cuyas combinaciones convexas producen todo el conjunto Θ de los diseños.

Se trata ahora de elegir una cantidad suficiente, r , de estos diseños iniciales o puntos, y luego explicitar ese conjunto formado por r diseños:

- el número suficiente viene dado por el teorema de Caratheodory
- la elección de los diseños constituye el algoritmo que presentamos.

El teorema de Caratheodory asegura la existencia de un óptimo con cardinal

$$C \leq \frac{m(m+1)}{2} + 1$$

(siendo m el número de parámetros del modelo). En el caso de criterios estrictamente decrecientes) se puede reducir a un cardinal

$$C \leq \frac{m(m+1)}{2},$$

como ya se ha visto. Llamaremos entonces

$$r = \frac{m(m+1)}{2} + 1$$

en el caso general, o

$$r = \frac{m(m+1)}{2}$$

si trabajamos con criterios estrictamente decrecientes.

El algoritmo consiste ahora en la elección (en principio arbitraria) de un conjunto de $r + 1$ diseños unipuntuales: “conjunto base”. Una vez elegido,

se calcula el óptimo del cierre convexo de estos diseños. Por el teorema de Caratheorory, existe un óptimo constituido solamente por la combinación convexa de r diseños, por lo que al menos uno de nuestros diseños iniciales no participaría en su creación. De hecho existiría un óptimo para este cierre convexo que tiene peso cero para al menos uno de estos diseños. El siguiente paso sería la eliminación de este diseño del conjunto base, y su sustitución por otro diseño, con lo que obtendríamos un nuevo conjunto base al que aplicar el mismo procedimiento.

Este algoritmo acabaría en $(Card(X) - r)$ pasos.

Diversas mejoras

Aunque en un principio hablamos de la elección arbitraria tanto de los diseños iniciales como del que introducimos en cada etapa del proceso, el hecho de haber comprobado empíricamente que el orden de elección de los diseños unipuntuales que se van incorporando en cada etapa puede tener una importancia trascendental, nos lleva a intentar mejorar el algoritmo introduciendo un orden más “racional” en la elección de los puntos (en el caso de que X esté ordenado).

Así el algoritmo tomaría la siguiente forma:

1. En principio se coge como conjunto base inicial el constituido por los

$$\left[\frac{r+2}{2} \right] = \left[\frac{r}{2} \right] + 1$$

primeros diseños unipuntuales de X ($[\cdot]$ significa “parte entera de...”), y los

$$\left[\frac{r+1}{2} \right]$$

últimos.

2. A continuación hallamos el ϕ -óptimo de las combinaciones convexas de los diseños de este conjunto base.
3. Eliminamos del conjunto base el diseño correspondiente al punto que participe con menor peso en la creación de este óptimo.

4. En la siguiente etapa se introduce el diseño siguiente a los primeros o el anterior a los últimos, según que el diseño eliminado sea de los primeros o los últimos respectivamente. Formamos así un nuevo conjunto base, y volvemos al paso 2.

Este procedimiento se basa en el hecho de que normalmente el óptimo “exige” probar el valor del modelo en los extremos del conjunto X , y luego, una vez asegurados los extremos (pueden ser los dos, o solamente uno si el valor del modelo en el otro está perfectamente claro), se apoya en algunos de los puntos interiores. Si ocurriera que se fueran eliminando solamente puntos del principio, o sólo del final, una variante del método sería comenzar a incorporar los diseños por el extremo opuesto: posiblemente estaríamos en el caso antes mencionado, aquél en que el valor en uno de los extremos queda completamente determinado por el modelo, y por tanto no necesitamos muestrear en ese extremo. Veamos la aplicación a un caso concreto.

Ejemplo 3.1 *Vamos a aplicar el algoritmo presentado a la obtención del diseño óptimo para el modelo*

$$y(x) = \alpha x^3 + \beta x^4 e^{-x^3}$$

siendo

$$X = \{0, 0.1, \dots, 0.9, 1\}$$

Como $m = 2$ (número de parámetros), $r = m(m+1)/2 = 3$. Luego existe un óptimo formado únicamente por 2 ó 3 puntos de X . Para empezar cogeremos pues 4 puntos de X que formarán el primer conjunto base. Según el orden prefijado, estos puntos son 0, 0.1, 0.9 y 1. Estas serían las etapas del algoritmo:

- *Calculamos el óptimo de las combinaciones convexas de los 4 puntos. Resulta el diseño*

$$\left\{ \begin{array}{cccc} 0 & 0.1 & 0.9 & 1 \\ 0 & 0 & 1/2 & 1/2 \end{array} \right\}$$

Por tanto eliminamos los puntos 0 y 0.1. Como los eliminados son ambos del principio, habrá que sustituirlos por otros dos puntos del principio: 0.2 y 0.3. Tenemos así un nuevo conjunto base.

- *Hallamos el óptimo. Obtenemos el diseño*

$$\left\{ \begin{array}{cccc} 0.2 & 0.3 & 0.9 & 1 \\ 0 & 0 & 1/2 & 1/2 \end{array} \right\}$$

Eliminamos los puntos 0.2 y 0.3. Los sustituimos por 0.4 y 0.5. Se forma el nuevo conjunto base.

- *Hallamos el óptimo. Obtenemos el diseño*

$$\left\{ \begin{array}{cccc} 0.4 & 0.5 & 0.9 & 1 \\ 0 & 0 & 1/2 & 1/2 \end{array} \right\}$$

Eliminamos los puntos 0.4 y 0.5. Los sustituimos por 0.6 y 0.7. Se forma el nuevo conjunto base.

- *Calculamos el óptimo. Obtenemos el diseño*

$$\left\{ \begin{array}{cccc} 0.6 & 0.7 & 0.9 & 1 \\ 0 & 0 & 1/2 & 1/2 \end{array} \right\}$$

Eliminamos los puntos 0.6 y 0.7. Sólo queda un punto: 0.8. Para formar el último conjunto base eliminamos el 0.6 (daría igual eliminar el 0.7) y añadimos el 0.8.

- *Hallamos el óptimo. El diseño óptimo definitivo será*

$$\left\{ \begin{array}{cccc} 0.8 & 0.7 & 0.9 & 1 \\ 1/2 & 0 & 0 & 1/2 \end{array} \right\}$$

Por lo tanto el algoritmo nos ha conducido al diseño

$$\left\{ \begin{array}{cc} 0.8 & 1 \\ 1/2 & 1/2 \end{array} \right\}$$

Obsérvese que el óptimo “desprecia” uno de los extremos: el 0. En realidad no necesitamos en absoluto probar en ese punto, ya que el modelo va a tomar siempre el valor 0 en él. En este caso podríamos haber empleado la variante del método general: en el primer paso, al eliminar los dos primeros puntos, incorporaríamos los dos inmediatamente anteriores a los últimos, 0.7 y 0.8. En el siguiente paso ya encontraríamos el óptimo.

Este método se puede intentar combinar con la información que podamos poseer para obtener más rápidamente los diseños. Por ejemplo si sabemos con bastante seguridad algunos de los puntos o pesos que intervienen en el óptimo, si existe simetría en el diseño, etc.

En algunos de los programas implementados con el ordenador se ha utilizado una correspondencia entre diseños con soporte en un conjunto determinado de $n + 1$ puntos (para los que queremos hallar sus pesos óptimos) y el compacto $[0, 1]^n$. Esta identificación es también original del autor, y a continuación la detallaremos, explicando brevemente sus principales propiedades:

Definición 3.1 *Se define*

$$\Upsilon_{x_1, \dots, x_{n+1}} : [0, 1]^n \longrightarrow \{\text{Diseños con soporte en } x_1, \dots, x_{n+1}\}$$

como la correspondencia

$$\alpha = \{\alpha_1, \dots, \alpha_n\} \longmapsto \left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_{n+1} \\ p_1 & p_2 & \cdots & p_{n+1} \end{array} \right\},$$

donde $p_1 = \alpha_1$, $p_i = \alpha_i(1 - \alpha_{i-1}) \cdots (1 - \alpha_1)$ para $i = 2, \dots, n$ y $p_{n+1} = (1 - \alpha_n) \cdots (1 - \alpha_1)$.

Teorema 3.1 *La correspondencia Υ (mientras no haya posibilidad de confusión se evitarán los subíndices) verifica las siguientes propiedades:*

1. *Está bien definida*
2. *Es epiyectiva*
3. *Su restricción a $[0, 1]^n$ es inyectiva*

Demostración: Para simplificar la notación nos referiremos al diseño

$$\xi = \left\{ \begin{array}{ccc} x_1 & \cdots & x_{n+1} \\ p_1 & \cdots & p_{n+1} \end{array} \right\}$$

simplemente como $\{p_1, \dots, p_{n+1}\}$.

1. Hay que probar que $\Upsilon(\alpha)$ es un diseño, es decir, que $\sum_{i=1}^{n+1} p_i = 1$. Pero si definimos

$$S_j = \alpha_j + \alpha_{j+1}[(1 - \alpha_j) + \dots \\ \dots + \alpha_n(1 - \alpha_{n-1}) \cdots (1 - \alpha_j) + (1 - \alpha_n) \cdots (1 - \alpha_j)]$$

se verifica $\sum_{i=1}^{n+1} p_i = S_1$ y $S_j = \alpha_j + (1 - \alpha_j)S_{j+1}$. Por tanto para acabar basta ver que $S_j = 1$, $j = 1, \dots, n$, que por la fórmula recurrente se reduce a la comprobación evidente de que $S_n = 1$.

2. Una antiimagen de $\{p_1, \dots, p_{n+1}\}$ es $\{\alpha_1, \dots, \alpha_n\}$, con

$$\alpha_1 = p_1; \quad \alpha_i = \left\{ \begin{array}{ll} 0 & \text{si } p_i = 0 \\ \frac{p_i}{1-p_1-\dots-p_{i-1}} & \text{si } p_i \neq 0 \end{array} \right\}, \quad i = 2, \dots, n$$

3. La inyección se comprueba trivialmente al restringirse a $[0, 1]^n$. \square

Observación 3.2 *Los únicos diseños que rompían la inyectividad tenían la forma,*

$$\left\{ \begin{array}{cccccc} x_1 & \cdots & x_i & x_{i+1} & \cdots & x_{n+1} \\ p_1 & \cdots & p_i & 0 & \cdots & 0 \end{array} \right\},$$

cuya antiimagen son los puntos

$$\left(p_1, \frac{p_2}{1-p_1}, \dots, \frac{p_i}{1-p_1-\dots-p_{i-1}}, 1, 0, \dots, 0 \right).$$

La restricción que proporciona la inyectividad estropea a su vez la epiyección, ya que ahora en el conjunto imagen todos los puntos tendrán pesos diferentes de cero.

La ventaja que se obtiene al utilizar esta correspondencia es la derivada de no tener que preocuparse de la restricción $\sum p_i = 1$, además de reducir la dimensión del conjunto de búsqueda.

Como se ha venido observando a lo largo de este capítulo, sería fundamental la definición de un orden en el conjunto de los diseños (o al menos en el de las matrices de información) que nos sirviera como base a la hora de elegir los diseños iniciales, o los que vamos incorporando en cada paso. Con este fin se estudió el tema, y se obtuvieron algunos resultados que se presentan en el capítulo siguiente.

Capítulo 4

Órdenes de las matrices de información. Relación con las funciones criterio

En este capítulo definiremos un orden en el conjunto de los diseños a partir del orden entre sus matrices de información asociadas. Introduciremos el concepto de comparabilidad de dos diseños. Analizaremos entonces los diseños más básicos, principales candidatos para diseños iniciales en los algoritmos que hemos propuesto en el capítulo anterior.

4.1 Orden de Loewner

Definición 4.1 Sean M y N dos matrices simétricas. Diremos que $N < M$ si $M - N$ es definido positiva; y que $N \leq M$ si $M - N$ es semidefinido positiva. La relación “ \leq ” establece un orden parcial en el conjunto de las matrices simétricas, llamado orden de Loewner.

Definición 4.2 Sean ξ y η dos diseños. Diremos que ξ y η son equivalentes, $\xi \sim \eta$, si y sólo si $M(\xi) = M(\eta)$. El diseño ξ no es peor que η , $\xi \preceq \eta$, si $M(\eta) \leq M(\xi)$. Finalmente diremos que ξ es mejor que η , $\xi \prec \eta$, si $M(\eta) < M(\xi)$.

Observación 4.1 *La definición que se acaba de dar es la más conveniente para la introducción del concepto de diseños comparables, que se hará a continuación. A pesar de que la definición usual se realiza a partir de la comparación de las varianzas generalizadas para todos los funcionales lineales, es equivalente a la dada cuando los diseños son no singulares (Proposición 2.6). Hay que hacer notar que $\prec \cup \sim$ no es lo mismo que \preceq .*

Definición 4.3 *Diremos que ξ y η son comparables respecto de la relación “ \prec ” si $M(\xi) < M(\eta)$ ó $M(\eta) < M(\xi)$. Serán comparables respecto de la relación “ \preceq ” si $M(\xi) \leq M(\eta)$ ó $M(\eta) \leq M(\xi)$. En general diremos que los diseños ξ, η son comparables respecto a una relación cualquiera “ \triangleleft ” si se verifica $M(\xi) \triangleleft M(\eta)$ ó $M(\eta) \triangleleft M(\xi)$.*

4.2 Propiedades y resultados sobre el orden de Loewner

En lo que sigue X será el conjunto de diseño, que supondremos incluido en \mathbb{R} . Partiremos del modelo lineal y como siempre supondremos la varianza de las observaciones igual a 1.

4.2.1 Matrices NND y orden de Loewner

Sea $N \in NND(m)$ = Matrices simétricas semidefinido positivas de orden m (nonnegative definite).

Se definen los conjuntos

$$C_p(N) = \{M \in NND(m) : N \geq M\}$$

$$C_g(N) = \{M \in NND(m) : M \geq N\}$$

Estudiemos algunas de sus propiedades:

- $C_p(N) \cap C_g(N) = \{N\}$

- $0 \in C_p(N)$
- Si $M \in C_p(N)$ y $a \leq 1$ entonces $aM \in C_p(N)$
- Si $M \in C_g(N)$ y $a \geq 1$ entonces $aM \in C_g(N)$
- $C_p(N)$ es convexo:

$$\begin{aligned} T, M \in C_p(N) &\Rightarrow N \geq T, N \geq M \Rightarrow \\ &\Rightarrow N - [\alpha T + (1 - \alpha) M] = \alpha [N - T] + (1 - \alpha) [N - M] \geq 0 \end{aligned}$$

- $C_g(N)$ es convexo:

$$\begin{aligned} T, M \in C_g(N) &\Rightarrow T \geq N, M \geq N \Rightarrow \\ &\Rightarrow [\alpha T + (1 - \alpha) M] - N = \alpha [T - N] + (1 - \alpha) [M - N] \geq 0 \end{aligned}$$

- $C_g(N) = N + NND(m)$

Observación 4.2 $NND(m)$ es un cono en el espacio de las matrices simétricas con vértice en el origen. Por tanto $C_g(N)$ será un cono de vértice N .

4.2.2 Relación “ \prec ”

Proposición 4.1 Los diseños unipuntuales no son comparables respecto a la relación “ \prec ”.

Demostración: Sean

$$\xi_{x_1} = \left\{ \begin{array}{c} x_1 \\ 1 \end{array} \right\}, \quad \xi_{x_2} = \left\{ \begin{array}{c} x_2 \\ 1 \end{array} \right\},$$

dos diseños unipuntuales. Se tiene que

$$M = M(\xi_{x_1}) - M(\xi_{x_2}) = (f_i(x_1)f_j(x_1) - f_i(x_2)f_j(x_2))_{i,j=1,\dots,m}.$$

Para que M fuera definido positiva sería necesario que el segundo determinante principal, D_2 , fuera positivo, pero

$$D_2 = -[f_1(x_1)f_2(x_2) - f_1(x_2)f_2(x_1)]^2 \leq 0$$

y por tanto ξ_{x_1}, ξ_{x_2} no pueden ser comparables. □

Corolario 4.1 *Los diseños de la forma*

$$\xi_\alpha = \left\{ \begin{array}{cc} x_1 & x_2 \\ \alpha & (1 - \alpha) \end{array} \right\}, \text{ y } \xi_\beta = \left\{ \begin{array}{cc} x_1 & x_2 \\ \beta & (1 - \beta) \end{array} \right\}$$

no son comparables respecto a la relación “ \prec ”.

Demostración: La matriz diferencia puede ponerse como

$$M(\xi_\alpha) - M(\xi_\beta) = (\alpha - \beta) [M(\xi_{x_1}) - M(\xi_{x_2})],$$

y por la Proposición 4.1 se concluye que esta matriz no puede ser definida positiva. \square

Proposición 4.2 *Los diseños ξ, η , tales que*

$$\text{Card}(X_\xi \cup X_\eta) < m$$

no son comparables respecto a la relación “ \prec ”.

Demostración: Sea $X_\xi \cup X_\eta = \{x_1, \dots, x_r\}$, con $r < m$.

$$\det(M(\xi) - M(\eta)) = \sum_{1 \leq i_1 \leq \dots \leq i_m \leq r} c_{i_1} \cdots c_{i_m} \underbrace{\det \left(f_t(x_{i_j}) \right)_{t,j=1,\dots,m}^2}_{D(i_1, \dots, i_m)^2}$$

donde $c_{i_j} = \xi(x_{i_j}) - \eta(x_{i_j})$ ó $(\xi(x_{i_j}) - \eta(x_{i_j}))$; y los determinantes $D(i_1, \dots, i_m)$ se anulan, ya que por hipótesis de partida al menos dos de las filas van a ser iguales. \square

Corolario 4.2 *Para i par, $i \leq m$, los diseños*

$$\xi_{x_i} = \left\{ \begin{array}{c} x_i \\ 1 \end{array} \right\}, \quad \xi_2 = \left\{ \begin{array}{ccc} x_1 & \cdots & x_{i-1} \\ \alpha_1 & \cdots & \alpha_{i-1} \end{array} \right\}$$

no son comparables respecto a la relación “ \prec ”. Para i impar, $i \leq m$, no puede ser $\xi_2 \prec \xi_{x_i}$.

Demostración: Supongamos i par. Si $i < m$ acabamos por la proposición anterior. Si $i = m$

$$\det(M(\xi_{x_i}) - M(\xi_2)) = (-1)^{i-1} \alpha_1 \cdots \alpha_{i-1} D(1, \dots, i)^2 \leq 0.$$

Además, sea i par o impar se tiene

$$\det(M(\xi_2) - M(\xi_{x_i})) = (-1) \alpha_1 \cdots \alpha_{i-1} D(1, \dots, i)^2 \leq 0,$$

lo que demuestra el enunciado. \square

Todos estos resultados indican la dificultad de establecer una prelación entre los diseños, especialmente entre los unipuntuales, como pretendíamos a la hora de poner en práctica el Algoritmo Sistemático descrito en el capítulo anterior. Al menos parece evidente esta complicación para la relación \prec . Examinamos ahora la relación \preceq .

4.2.3 Relación “ \preceq ”

Teorema 4.1 (*Caracterización geométrica del orden parcial de los diseños unipuntuales*)

Dados $x_1, x_2 \in X$, se tiene

$$\xi_{x_1} \preceq \xi_{x_2} \iff M(\xi_{x_1}) = k^2 M(\xi_{x_2}), \quad k > 1$$

Demostración: Supongamos $\xi_{x_1} \preceq \xi_{x_2}$. Entonces los menores principales tendrán la forma siguiente:

$$D_l = \det(f_i(x_1)f_j(x_1) - f_i(x_2)f_j(x_2))_{i,j=1,\dots,l} \geq 0, \quad l = 1, \dots, m$$

que nos dice:

- $D_1 \geq 0 \Rightarrow f_1(x_1^2) \geq f_1(x_2)^2$
- $D_2 \geq 0$. Pero

$$D_2 = - \begin{vmatrix} f_1(x_1) & f_2(x_1) \\ f_1(x_2) & f_2(x_2) \end{vmatrix}^2 \leq 0.$$

La única posibilidad de igualdad es que la segunda fila sea proporcional a la primera: $f_1(x_2) = c f_1(x_1)$ y $f_2(x_2) = c f_2(x_1)$, con $c^2 \leq 1$ por la condición anterior. Pero como una matriz semidefinido positiva lo

sigue siendo por transformaciones lineales, este resultado se mantiene si en el modelo intercambiamos $f_i(x)$ por $f_j(x)$. Así tenemos $f_i(x_2) = c f_i(x_1) \quad \forall i = 2, \dots, m$, siendo c la misma para todos. Por tanto

$$M(\xi_{x_1}) = k^2 M(\xi_{x_2})$$

con $k^2 = (1/c)^2 \geq 1$.

Si, recíprocamente, suponemos $M(\xi_{x_1}) = k^2 M(\xi_{x_2})$, con $k > 1$, entonces

$$M(\xi_{x_1}) - M(\xi_{x_2}) = (k^2 - 1)M(\xi_{x_2}) \geq 0$$

por ser $M(\xi_{x_2})$ una matriz de información. □

Proposición 4.3 Sean $x_1, x_2 \in X$, y consideremos los diseños

$$\xi_\alpha = \left\{ \begin{array}{cc} x_1 & x_2 \\ \alpha & (1 - \alpha) \end{array} \right\} \text{ y } \xi_\beta = \left\{ \begin{array}{cc} x_1 & x_2 \\ \beta & (1 - \beta) \end{array} \right\}, \quad \alpha \geq \beta.$$

Entonces se cumple:

$$\xi_\alpha \preceq \xi_\beta \Leftrightarrow M(\xi_{x_2}) \leq M(\xi_{x_1})$$

Demostración: Se tiene

$$M = M(\xi_\alpha) - M(\xi_\beta) = (\alpha - \beta)[M(\xi_{x_1}) - M(\xi_{x_2})]$$

□

Proposición 4.4 Consideremos el modelo lineal con dos parámetros y $x_1, x_2, x_3 \in X$, siendo ξ_{x_1} y ξ_{x_2} no comparables entre sí (" \preceq "). Consideremos los diseños

$$\xi_{x_3} = \left\{ \begin{array}{c} x_3 \\ 1 \end{array} \right\} \text{ y } \xi_\alpha = \left\{ \begin{array}{cc} x_1 & x_2 \\ \alpha & (1 - \alpha) \end{array} \right\}$$

y denotemos

$$D_{ij} = \left| \begin{array}{cc} f_1(x_i) & f_2(x_i) \\ f_1(x_j) & f_2(x_j) \end{array} \right|$$

Si $1/2 \leq \beta < \alpha$ y $D_{13}^2 \geq D_{23}^2$, entonces $\xi_{x_3} \preceq \xi_\alpha$ implica $\xi_{x_3} \preceq \xi_\beta$.

Demostración: Ardanuy y otros (1999) presentaron un desarrollo del determinante que es válido no sólo para diseños, sino para cualquier matriz asociada a una medida (por ejemplo, para una combinación lineal de diseños, aunque no sea diseño). A este respecto puede consultarse la Proposición 2.2. En nuestro caso, tenemos

$$\det [M(\xi_{x_3}) - M(\xi_\alpha)] = -\alpha D_{13}^2 - (1 - \alpha)D_{23}^2 + \alpha(1 - \alpha)D_{12}^2,$$

y por tanto el hecho de que, $M(\xi_{x_3}) - M(\xi_\alpha) \geq 0$ implica que $D_{12}^2 \geq D_{13}^2$ y $D_{12}^2 \geq D_{23}^2$. Entonces

$$\begin{aligned} \det [M(\xi_{x_3}) - M(\xi_\beta)] &= -\beta D_{13}^2 - (1 - \beta)D_{23}^2 + \beta(1 - \beta)D_{12}^2 \\ &> -\alpha D_{13}^2 - (1 - \alpha)D_{23}^2 + \alpha(1 - \alpha)D_{12}^2 \\ &= \det [M(\xi_{x_3}) - M(\xi_\alpha)] \geq 0 \end{aligned}$$

ya que $\beta(1 - \beta) > \alpha(1 - \alpha)$ y

$$-\beta D_{13}^2 - (1 - \beta)D_{23}^2 - [-\alpha D_{13}^2 - (1 - \alpha)D_{23}^2] = (\alpha - \beta)(D_{13}^2 - D_{23}^2) > 0.$$

Faltaría ver que

$$f_1(x_3)^2 - \beta f_1(x_1)^2 - (1 - \beta)f_1(x_2)^2 \geq 0,$$

lo cual ha de ser cierto, porque de no ser así tendríamos $M(\xi_\beta) \geq M(\xi_{x_3})$, luego $M(\xi_\beta) \geq M(\xi_\alpha)$. Esto implicaría (ver Proposición 4.3) que $M(\xi_{x_2}) \geq M(\xi_{x_1})$, lo que contradice la hipótesis de partida. \square

Observación 4.3 Siguiendo con el modelo biparamétrico, sean $x_1, x_2, x_3 \in X \subset \mathbb{R}$,

$$f(x) = (1, x)^T, \quad \xi_\alpha = \left\{ \begin{array}{cc} x_1 & x_2 \\ \alpha & (1 - \alpha) \end{array} \right\}, \quad M = M(\xi_{x_3}) - M(\xi_\alpha).$$

Supongamos además que ξ_{x_1} y ξ_{x_2} son no comparables respecto de \preceq . Veamos que en caso de que ξ_{x_3} y ξ_α lo sean, ξ_α será siempre mejor que ξ_{x_3} :

Por ser comparables ha de verificarse $\det(M) \geq 0$. Además, por la demostración de la Proposición 4.4 ha de ser $D_{12}^2 \geq D_{13}^2$ y $D_{12}^2 \geq D_{23}^2$, que en nuestro caso significa

$$|x_1 - x_2| \geq |x_1 - x_3| \text{ y } |x_1 - x_2| \geq |x_2 - x_3|,$$

de donde se deduce que:

$$x_3 = \beta x_1 + (1 - \beta)x_2$$

con lo que será $\det(M) = -(\alpha - \beta)^2(x_2 - x_1)^2 \leq 0$. Esto, por la hipótesis $M \geq 0$, implica la igualdad, cosa que sólo es posible (excluyendo el caso $x_1 = x_2$) si $\alpha = \beta$. Y con este resultado se obtiene $M(\xi_\alpha) \geq M(\xi_{x_3})$.

El resultado nos viene a decir que dentro de un conjunto de diseños comparables un diseño unipuntual nunca es mejor que uno con dos puntos. Esto no es sorprendente, ya que los criterios globales nos conducen a diseños con al menos m puntos en su soporte.

Capítulo 5

Criterios característicos

5.1 Introducción y definiciones

Dos de los criterios más conocidos en diseño óptimo de experimentos son los del determinante y la traza de la matriz de información. En este capítulo presentaremos un conjunto general de criterios basados en los coeficientes del polinomio característico de esta matriz, que podríamos decir estarían situados entre los dos criterios antes mencionados, siendo éstos, casos particulares. Mientras que la A -optimización considera la media de las varianzas, la D -optimización tiene en cuenta todas las varianzas junto con las covarianzas y hace mínimo el volumen del elipsoide de confianza de los parámetros. De forma análoga se podría buscar interpretaciones al resto de los coeficientes. Así, por ejemplo, el segundo tiene una gran relación con los pares de coeficientes de regresión, como veremos más adelante.

Muchos de los resultados desarrollados en este capítulo pueden consultarse en López Fidalgo y Rodríguez Díaz (1998).

5.1.1 Funciones simétricas elementales

Sea M simétrica y semidefinido positiva de orden m . Entonces el polinomio característico se puede poner como

$$\begin{aligned} P_c(M) &= |x I - M| \\ &= \varphi_0(M) x^m - \varphi_1(M) x^{m-1} + \varphi_2(M) x^{m-2} + \dots + (-1)^m \varphi_m(M), \end{aligned}$$

donde

$$\begin{aligned}\varphi_k(M) &= \sum_{i_1 < \dots < i_k} \lambda_{i_1} \dots \lambda_{i_k} \quad 1 \leq k \leq m \\ \varphi_0(M) &= 1,\end{aligned}\tag{5.1}$$

son las funciones simétricas en los valores propios de M , $\lambda_1, \dots, \lambda_m$. Los coeficientes primero y último son precisamente

$$\varphi_1(M) = \text{tr}(M) \quad \text{y} \quad \varphi_m(M) = \det(M),$$

que dan lugar a los conocidos criterios de A y D -optimización:

$$\Phi_A[M(\xi)] = \text{tr}[M^{-1}(\xi)]; \quad \Phi_D[M(\xi)] = \log \det[M^{-1}(\xi)].$$

Las funciones $\varphi_1, \dots, \varphi_m$ se llaman *funciones simétricas elementales* de la matriz o *coeficientes característicos*, y la caracterizan como función lineal.

Notación 5.1 Dada una matriz $m \times m$ M , denotaremos $[M]_{i_1, \dots, i_k}$ ($1 \leq k \leq m$, $1 \leq i_1 < \dots < i_k \leq m$) a la submatriz principal de M formada por los elementos situados en las filas y columnas i_1, \dots, i_k . En particular, $[M]_{1, \dots, k}$ será denotada simplemente por $[M]_k$.

Observación 5.1 En ocasiones la matriz que nos ocupa no está completamente definida, sino que depende del valor de algún parámetro. Entonces es mejor calcular los coeficientes característicos de la manera siguiente:

$$\varphi_k(M) = \sum_{i_1 < \dots < i_k} \det [M_{i_1, \dots, i_k}] \quad 1 \leq k \leq m.\tag{5.2}$$

5.1.2 Propiedades

Todas las matrices que aparezcan de ahora en adelante se considerarán simétricas y semidefinido positivas, lo cual es una condición más fuerte de lo que necesitamos, ya que evidentemente nuestro interés se centra en las matrices de información, y dentro de éstas, en las regulares, para evitar los casos más particulares de diseños singulares. Primeramente hacemos observar que para todos los cálculos vamos a poder suponer que M es una matriz diagonal con

elementos $\lambda_1, \dots, \lambda_m$, porque el polinomio característico de M es invariante por cambios de base de la forma $A^{-1}MA$, siendo A no singular. Por tanto, a partir de ahora será $M = \text{diag}(\lambda_1, \dots, \lambda_m)$, donde $\lambda_1, \dots, \lambda_m$ son los valores propios de M .

Por el mismo motivo tenemos la propiedad circular

$$\varphi_k(MN) = \varphi_k(NM),$$

para cualquier matriz regular M y cualquier matriz cuadrada N .

De la misma manera, de ahora en adelante cuando manejemos varias matrices simétricas y semidefinido positivas podremos suponer que una es diagonal y el resto simétricas, ya que dada una matriz simétrica siempre existe una base ortogonal en la que diagonaliza, y que por ser ortogonal transforma matrices simétricas en simétricas y semidefinido positivas en semidefinido positivas.

Proposición 5.1 *Se cumple*

$$\varphi_k(\lambda M) = \lambda^k \varphi_k(M)$$

Demostración: Se deduce fácilmente de (5.2), ya que

$$\det[\lambda M]_{i_1, \dots, i_k} = \lambda^k \det[M_{i_1, \dots, i_k}]$$

□

Recordemos el siguiente resultado:

Proposición 5.2 (*Desigualdad de Hadamard*) *Si N es una matriz semidefinido positiva se verifica*

$$\det N \leq \prod_{i=1}^m n_{ii}.$$

Proposición 5.3 *Las funciones $\varphi_k(M)$ son crecientes respecto del orden de Loewner.*

Demostración: Sean M y N , simétricas y semidefinido positivas, con $M \geq N$. Por las razones ya mencionadas podemos suponer que M es diagonal, $M = \text{diag}(\lambda_1, \dots, \lambda_m)$, y $N = (n_{ij})$ simétrica. La condición $M \geq N$ implica que $\lambda_i \geq n_{ii}$, para todo $i = 1, \dots, m$. Esto es claro para $i = 1$, e igualmente cierto para el resto debido a que mediante un adecuado cambio de base siempre podemos intercambiar λ_1 y λ_i , ocurriendo lo propio con n_{11} y n_{ii} y conservándose el mismo orden en las matrices resultantes. Por tanto

$$\begin{aligned} \varphi_k(M) &= \sum_{i_1 < \dots < i_k} \lambda_{i_1} \dots \lambda_{i_k} \\ &\geq \sum_{i_1 < \dots < i_k} n_{i_1 i_1} \dots n_{i_k i_k} \\ &\geq \sum_{i_1 < \dots < i_k} \begin{vmatrix} n_{i_1 i_1} & \dots & n_{i_1 i_k} \\ \vdots & \ddots & \vdots \\ n_{i_k i_1} & \dots & n_{i_k i_k} \end{vmatrix} = \varphi_k(N) \end{aligned}$$

donde el último paso es la desigualdad de Hadamard. \square

Si vamos a definir criterios basados en las funciones características, será interesante previamente relacionarlos entre ellos, y ver cómo actúan sobre la matriz inversa. Existe una relación fundamental entre las funciones características aplicadas a una matriz M y a su inversa:

Proposición 5.4 *Dada M definido positiva se verifica*

$$\varphi_k(M^{-1}) = \varphi_m(M^{-1}) \varphi_{m-k}(M) \quad (5.3)$$

Demostración: A partir de la Definición 5.1, y utilizando $M = \text{diag}(\lambda_1, \dots, \lambda_m)$ se tiene

$$\begin{aligned} \varphi_m(M^{-1}) \varphi_{m-k}(M) &= (\lambda_1^{-1} \dots \lambda_m^{-1}) \sum_{j_1 < \dots < j_{m-k}} \lambda_{j_1} \dots \lambda_{j_{m-k}} \\ &= \sum_{j_1 < \dots < j_{m-k}} \frac{\lambda_{j_1} \dots \lambda_{j_{m-k}}}{(\lambda_1^{-1} \dots \lambda_m^{-1})} \\ &= \sum_{i_1 < \dots < i_k} \frac{1}{\lambda_{i_1} \dots \lambda_{i_k}} = \varphi_k(M^{-1}) \end{aligned}$$

\square

Observación 5.2 *Nótese la importancia de esta sencilla relación: la inmensa mayoría de las funciones criterio actúan sobre la inversa de la matriz de información del diseño. La obtención de esta inversa en cada etapa de un proceso algorítmico se convierte en una de las tareas más costosas, computacionalmente hablando. Sin embargo, el uso de la fórmula consigue evitar este engorroso cálculo, ya que el determinante de M^{-1} ($\varphi_m(M^{-1})$) es simplemente el inverso del determinante de la matriz M . Esto se aplica a todos los coeficientes característicos, incluido el primero y muy utilizado, A-optimización. Las funciones $\varphi_{m-k}(M)$ no presentan problemas para su obtención en ningún caso, utilizando (5.2).*

Corolario 5.1 *Si m es par y $k = m/2$ se tiene:*

$$\frac{\varphi_k(M)}{\varphi_k(M^{-1})} = \det(M)$$

Como procedimiento de recurrencia será interesante obtener $\varphi_k(M)$ a partir de $\varphi_1(M), \dots, \varphi_{k-1}(M)$. Los siguientes resultados se dirigen a buscar dicha relación.

Lema 5.1 *Dado k entero positivo*

$$\sum_{t_1 < \dots < t_{k-j}; t \neq t_r} \lambda_{t_1} \dots \lambda_{t_{k-j}} \lambda_t^j = \sum_{i=j}^k (-1)^{i+j} \varphi_{k-i}(M) \varphi_1(M^i) \quad (5.4)$$

donde $1 \leq j \leq k$.

Demostración: Se demostrará por inducción descendente sobre j :

Para $j = k$ el miembro de la izquierda se reduce a $\sum_{t=1}^m \lambda_t^k$, y el de la derecha es

$$(-1)^{2k} \varphi_0(M) \varphi_1(M^k) = \text{tr}(M^k) = \sum_{t=1}^m \lambda_t^k$$

Para $j = k - 1$

$$\begin{aligned} \sum_{t \neq t_1} \lambda_{t_1} \lambda_t^{k-1} &= \left(\sum_{t_1=1}^m \lambda_{t_1} \right) \left(\sum_{t=1}^m \lambda_t^{k-1} \right) - \sum_{t=1}^m \lambda_t^k \\ &= \varphi_1(M) \varphi_1(M^{k-1}) - \varphi_1(M^k). \end{aligned}$$

Para $j < k - 1$

$$\begin{aligned} & \sum_{t_1 < \dots < t_{k-j}} \sum_{t \neq t_r} \lambda_{t_1} \dots \lambda_{t_{k-j}} \lambda_t^j \\ &= \left(\sum_{t_1 < \dots < t_{k-j}} \lambda_{t_1} \dots \lambda_{t_{k-j}} \right) \left(\sum_{t=1}^m \lambda_t^j \right) - \sum_{t_1 < \dots < t_{k-j-1}} \sum_{t \neq t_r} \lambda_{t_1} \dots \lambda_{t_{k-j-1}} \lambda_t^{j+1} \end{aligned}$$

que por hipótesis de inducción, será

$$\begin{aligned} & \varphi_{k-j}(M) \varphi_1(M^j) - \sum_{i=j+1}^k (-1)^{i+j+1} \varphi_{k-i}(M) \varphi_1(M^i) \\ &= \sum_{i=j}^k (-1)^{i+j} \varphi_{k-i}(M) \varphi_1(M^i), \end{aligned}$$

lo que concluye la demostración. \square

Teorema 5.1 *Se tiene la igualdad*

$$\varphi_k(M) = \frac{1}{k} \sum_{i=1}^k (-1)^{i-1} \varphi_{k-i}(M) \varphi_1(M^i) \quad (5.5)$$

Demostración: Podemos descomponer

$$\begin{aligned} \varphi_{k-1}(M) \varphi_1(M) &= \left(\sum_{t_1 < \dots < t_{k-1}} \lambda_{t_1} \dots \lambda_{t_{k-1}} \right) \left(\sum_{t=1}^m \lambda_t \right) \\ &= k \sum_{t_1 < \dots < t_k} \lambda_{t_1} \dots \lambda_{t_k} + \sum_{t_1 < \dots < t_{k-2}; t \neq t_j} \lambda_{t_1} \dots \lambda_{t_{k-2}} \lambda_t^2 \end{aligned}$$

y entonces

$$\varphi_k(M) = \frac{1}{k} \left[\varphi_{k-1}(M) \varphi_1(M) - \sum_{t_1 < \dots < t_{k-2}; t \neq t_j} \lambda_{t_1} \dots \lambda_{t_{k-2}} \lambda_t^2 \right].$$

Por el Lema 5.1 se concluye. \square

Observación 5.3 *Podría obtenerse $\varphi_k(M)$ en función de $\varphi_1(M^i)$, $i = 1, \dots, k$ sin más que aplicar la fórmula anterior recurrentemente para φ_{k-1} , φ_{k-2} , \dots , φ_2 .*

5.1.3 Funciones características

Estamos interesados en la creación de nuevas funciones criterio a partir de los coeficientes característicos. Especialmente importante será su aplicación a las inversas de las matrices de información, que como ya hemos visto son proporcionales a la matriz de covarianzas de las estimaciones de los parámetros.

Definición 5.1 *Llamaremos funciones características de una matriz M a los coeficientes característicos de su inversa. Esto es*

$$\Phi_{Ch_k}(M) = \varphi_k(M^{-1})$$

Con el objeto de probar la convexidad de estas funciones, propiedad siempre deseable en cualquier función criterio, veremos algunos resultados previos.

Proposición 5.5 *Para cualquier función real v definida sobre un convexo se verifica:*

1. Si v es cóncava, entonces es log-cóncava
2. Si v es log-convexa, entonces es convexa

Demostración:

1. Si v es cóncava se tiene

$$v[\alpha M + (1 - \alpha)N] \geq \alpha v(M) + (1 - \alpha)v(N),$$

que al aplicar el logaritmo, que conserva la desigualdad por ser creciente, produce

$$\begin{aligned} \log v[\alpha M + (1 - \alpha)N] &\geq \log[\alpha v(M) + (1 - \alpha)v(N)] \\ &\geq \alpha \log[v(M)] + (1 - \alpha) \log[v(N)], \end{aligned}$$

siendo la última desigualdad debida a la concavidad de la función log.

2. Si v no fuera convexa, existirían M , N y $0 < \alpha < 1$ tales que

$$v[\alpha M + (1 - \alpha)N] > \alpha v(M) + (1 - \alpha)v(N),$$

y del mismo modo que antes, al aplicar el logaritmo y usar que éste es creciente y cóncavo obtendríamos

$$\log v[\alpha M + (1 - \alpha)N] > \alpha \log v(M) + (1 - \alpha) \log v(N),$$

lo que contradice la hipótesis de partida. \square

Lema 5.2 Sean M y N matrices $m \times m$ definido positivas, y H una matriz $s \times m$ de rango s ($s \leq m$). Entonces se verifica

$$(HM^{-1}H^t)^{-1} + (HN^{-1}H^t)^{-1} \leq [H(M + N)^{-1}H^t]^{-1} \quad (5.6)$$

La demostración puede encontrarse en Pázman (1986), pg 102. En ella se utilizan varias desigualdades, entre ellas la de Schwarz.

Lema 5.3 La función

$$M \mapsto [M^{-1}]_{i_1, \dots, i_k}^{-1}$$

es cóncava, es decir, si M y N son matrices no singulares de rango m y $0 \leq \alpha \leq 1$ se cumple

$$[(\alpha M + (1 - \alpha)N)^{-1}]_{i_1, \dots, i_k}^{-1} \geq \alpha [M^{-1}]_{i_1, \dots, i_k}^{-1} + (1 - \alpha) [N^{-1}]_{i_1, \dots, i_k}^{-1} \quad (5.7)$$

Demostración: Sin pérdida de generalidad podemos suponer que $(i_1, \dots, i_k) = (1, \dots, k)$. Se puede escribir $[M]_k = HMH^t$, siendo H la matriz $k \times m$ definida por $H_{i,i} = 1$, $H_{i,j} = 0$ si $i \neq j$. Se cumple entonces

$$\begin{aligned} & \alpha [M^{-1}]_k^{-1} + (1 - \alpha) [N^{-1}]_k^{-1} \\ &= \alpha [HM^{-1}H^t]^{-1} + (1 - \alpha) [HN^{-1}H^t]^{-1} \\ &= [H\alpha M^{-1}H^t]^{-1} + [H(1 - \alpha)N^{-1}H^t]^{-1} \\ &\leq [H(\alpha M + (1 - \alpha)N)^{-1}H^t]^{-1} \\ &= [(\alpha M + (1 - \alpha)N)^{-1}]_k^{-1}, \end{aligned}$$

donde la desigualdad se deduce del Lema 5.6. \square

Teorema 5.2 *La función criterio $\Phi_{Ch_k}(M) = \varphi_k(M^{-1})$ es convexa, $k = 1, \dots, m$.*

Demostración: Podemos escribir

$$\Phi_{Ch_k}(M) = \sum_{1 \leq i_1 < \dots < i_k \leq m} \det [M^{-1}]_{i_1, \dots, i_k},$$

luego basta probar que las funciones

$$M \mapsto \det [M^{-1}]_{i_1, \dots, i_k}$$

son convexas, y por la Proposición 5.5 es suficiente ver la log-convexidad. De nuevo podemos suponer $(i_1, \dots, i_k) = (1, \dots, k)$, y entonces

$$\begin{aligned} & \log \det [(\alpha M + (1 - \alpha)N)^{-1}]_k \\ &= -\log \det [(\alpha M + (1 - \alpha)N)^{-1}]_k^{-1} \\ &\leq -\log \det \left(\alpha [M^{-1}]_k^{-1} + (1 - \alpha) [N^{-1}]_k^{-1} \right) \\ &\leq \alpha \left(-\log \det [M^{-1}]_k^{-1} \right) + (1 - \alpha) \left(-\log \det [N^{-1}]_k^{-1} \right) \\ &= \alpha \log \det [M^{-1}]_k + (1 - \alpha) \log \det [N^{-1}]_k, \end{aligned}$$

con lo que se acaba. La primera desigualdad viene de (5.7), teniendo en cuenta que el determinante conserva el orden de Loewner y el logaritmo es creciente. La segunda se deduce de la convexidad del criterio de D -optimización. \square

5.2 Continuidad y Diferenciabilidad

La continuidad y la diferenciabilidad de $\varphi_k(M)$ son inmediatas a partir de la definición. De hecho son polinomios homogéneos en las componentes de M . Centrémonos pues en el cálculo de su gradiente. Para ello usaremos el siguiente resultado:

Proposición 5.6 *El gradiente de $\varsigma_k(M) = \varphi_1(M^k)$ (M simétrica) es*

$$\nabla \varsigma_k(M) = k M^{k-1} \quad (5.8)$$

Demostración: Se tiene

$$\frac{\partial}{\partial m_{ij}} \varsigma_k(M) = \frac{\partial}{\partial m_{ij}} \sum_t \sum_{r_1, \dots, r_{k-1}} m_{tr_1} m_{r_1 r_2} \cdots m_{r_{k-1} t},$$

que se puede separar en tres casos según m_{ij} sea el primero, uno de los centrales o el último factor de cada sumando. Tras derivar quedaría

$$\begin{aligned} \frac{\partial}{\partial m_{ij}} \varsigma_k(M) &= \sum_{r_2, \dots, r_{k-1}} m_{j r_2} \cdots m_{r_{k-1} i} \\ &+ \sum_t \sum_{\substack{r_1, \dots, \widehat{r_s}, \widehat{r_{s+1}}, \dots, r_{k-1} \\ s=1, \dots, k-2}} m_{tr_1} \cdots m_{r_{s-1} i} m_{j r_{s+2}} \cdots m_{r_{k-1} t} \\ &+ \sum_{r_1, \dots, r_{k-2}} m_{j r_1} \cdots m_{r_{k-2} i} \end{aligned}$$

El segundo sumatorio se puede reordenar de la siguiente manera:

$$\sum_{s=1, \dots, k-2} \sum_{t, r_1, \dots, \widehat{r_s}, \widehat{r_{s+1}}, \dots, r_{k-1}} m_{j r_{s+2}} \cdots m_{r_{k-1} t} m_{tr_1} \cdots m_{r_{s-1} i},$$

y por tanto en todo el conjunto tenemos k veces cada sumatorio

$$\sum_{r_2, \dots, r_{k-1}} m_{j r_2} \cdots m_{r_{k-1} i} = (M^{k-1})_{ji},$$

el elemento (j, i) de la matriz M^{k-1} , lo que conduce al resultado final

$$\frac{\partial}{\partial m_{ij}} \varsigma_k(M) = k (M^{k-1})_{ji} = k (M^{k-1})_{ij}$$

□

Proposición 5.7 *El gradiente de los coeficientes característicos es*

$$\nabla \varphi_k(M) = \sum_{j=0}^{k-1} (-1)^j \varphi_{k-j-1}(M) M^j. \quad (5.9)$$

Demostración: Se probará por inducción. El caso $k = 1$ es trivial. Si $k > 1$, por (5.5) se tiene

$$\begin{aligned}\nabla\varphi_k(M) &= \frac{1}{k} \sum_{i=1}^k (-1)^{i-1} \nabla [\varphi_{k-i}(M) \varphi_1(M^i)] \\ &= \frac{1}{k} \sum_{i=1}^k (-1)^{i-1} [\nabla\varphi_{k-i}(M) \varphi_1(M^i) + \varphi_{k-i}(M) \nabla\varphi_1(M^i)]\end{aligned}$$

Usando la hipótesis de inducción y (5.8) resulta

$$\frac{1}{k} \sum_{i=1}^k (-1)^{i-1} \left[\sum_{j=0}^{k-i-1} (-1)^j \varphi_{k-i-j-1}(M) M^j \varphi_1(M^i) + \varphi_{k-i}(M) i M^{i-1} \right],$$

que tras reordenar sumandos y utilizar nuevamente (5.5) se puede poner como

$$\sum_{j=0}^{k-1} (-1)^j \varphi_{k-j-1}(M) M^j$$

□

Corolario 5.2 $\Phi_{Ch_k}(M) = \varphi_k(M^{-1})$ es una función criterio diferenciable cuyo gradiente es

$$\begin{aligned}\nabla\Phi_{Ch_k}(M) &= \sum_{i=k}^m \varphi_i(M^{-1}) (-M)^{i-k-1} \\ &= - \sum_{i=0}^{k-1} \varphi_i(M^{-1}) (-M)^{i-k-1}\end{aligned}$$

Demostración: Utilizando la expresión (5.3) el gradiente se puede poner como

$$\begin{aligned}\nabla\Phi_{Ch_k}(M) &= \nabla\varphi_k(M^{-1}) \\ &= \nabla [\varphi_m(M^{-1}) \varphi_{m-k}(M)] \\ &= \nabla\varphi_m(M^{-1}) \varphi_{m-k}(M) + \varphi_m(M^{-1}) \nabla\varphi_{m-k}(M)\end{aligned}$$

Ahora, por (5.9) y teniendo en cuenta que $\nabla \log \det(M) = M^{-1}$ (ver p.e. Pázman (1986) pg. 81) podemos obtener

$$\nabla\varphi_m(M^{-1}) = \nabla \log \varphi_m(M^{-1}) \varphi_m(M^{-1}) = -M^{-1} \varphi_m(M^{-1}),$$

lo que nos lleva finalmente a

$$\nabla \Phi_{Ch_k}(M^{-1}) = \sum_{i=k}^m \varphi_i(M^{-1}) (-M)^{i-k-1}.$$

Alternativamente podríamos expresarlo también como

$$\nabla \Phi_{Ch_k}(M^{-1}) = - \sum_{i=0}^{k-1} \varphi_i(M^{-1}) (-M)^{i-k-1}$$

utilizando que

$$M^{-(m-k-1)} \sum_{i=0}^m \varphi_i(M^{-1}) (-M)^{i-k-1} = \sum_{i=0}^m \varphi_i(M^{-1}) (-M)^{-(m-i)} = 0,$$

puesto que se trata del polinomio característico de la matriz M^{-1} aplicado a la propia matriz.

Así se puede utilizar según convenga cualquiera de las dos fórmulas, una (la primera) más adecuada para valores altos de k , y la otra mejor para valores pequeños de k . \square

Proposición 5.8 *Una condición necesaria y suficiente para que el diseño ξ^* sea Φ_{Ch_k} -óptimo es que se verifique*

$$\begin{aligned} \min \sum_{i=k}^m (-1)^{-k+i-1} \varphi_i[M^{-1}(\xi^*)] f^t(x) M^{-k+i-1}(\xi^*) f(x) \\ = \sum_{i=k}^m (-1)^{-k+i-1} \varphi_i[M^{-1}(\xi^*)] \varphi_1[M^{-k+i}(\xi^*)] \end{aligned}$$

Demostración: Por el Teorema 2.5 la condición que ha de cumplirse es

$$\min_{x \in X} f^t(x) \nabla \Phi_{Ch_k}[M(\xi^*)] f(x) = Tr M(\xi^*) \nabla \Phi_{Ch_k}[M(\xi^*)]$$

En este caso, utilizando el Corolario 5.2 se tiene

$$\begin{aligned} \min_{x \in X} f^t(x) \nabla \Phi_{Ch_k}[M(\xi^*)] f(x) \\ = \sum_{i=k}^m (-1)^{-k+i-1} \varphi_i[M^{-1}(\xi^*)] f^t(x) M^{-k+i-1}(\xi^*) f(x) \end{aligned}$$

y

$$\text{Tr}M(\xi^*) \nabla \Phi_{Ch_k}[M(\xi^*)] = \sum_{i=k}^m (-1)^{-k+i-1} \varphi_i[M^{-1}(\xi^*)] \varphi_1[M^{-k+i}(\xi^*)],$$

lo que prueba el enunciado. \square

Observación 5.4 *También hay que hacer notar que se podrían considerar funciones criterio más generales basadas en estos coeficientes, como*

$$\Phi_{p,k}(M) = \left[\frac{1}{\binom{m}{k}} \sum_{i_1 < \dots < i_k} (\lambda_{i_1} \dots \lambda_{i_k})^{-\frac{p}{k}} \right]^{\frac{1}{p}}.$$

La división y la raíz k -ésima se deben a que sumamos $\binom{m}{k}$ productos de k autovalores y por tanto así se considera un promedio. El resto se introduce como una generalización de los criterios Φ_p vistos en el Capítulo 2. En particular se podría expresar el criterio del determinante simplemente tomando $p = 1$ y $k = m$, que es equivalente al definido, por ser log una función creciente. El gradiente de estos criterios generales es

$$\nabla \Phi_{p,k}(M) = \frac{1}{k \binom{m}{k}^{\frac{1}{p}}} \varphi_k \left(M^{-\frac{p}{k}} \right)^{\frac{1-p}{p}} \sum_{i=k}^m \varphi_i \left(M^{-\frac{p}{k}} \right) (-1)^{i-k-1} M^{\frac{pi}{k}-p-1},$$

donde hemos utilizado la primera expresión dada para $\nabla \Phi_{Ch_k}(M)$.

Opcionalmente se podrían haber definido tomando la raíz k -ésima global, en vez de hacerlo para cada sumando:

$$\Phi_{p,k}(M) = \left[\frac{1}{\binom{m}{k}} \sum_{i_1 < \dots < i_k} (\lambda_{i_1} \dots \lambda_{i_k})^{-p} \right]^{\frac{1}{k p}}.$$

5.3 La segunda función característica

A continuación vamos a ver algunos resultados particulares para φ_2 . Primero veamos cómo actúa sobre la suma de matrices:

Proposición 5.9 *Dadas dos matrices simétricas y semidefinido positivas de orden m , M y N , se verifica*

$$\varphi_2(M + N) = \varphi_2(M) + \varphi_2(N) + [\varphi_1(M) \varphi_1(N) - \varphi_1(M N)] \quad (5.10)$$

Demostración: Podemos suponer $M = \text{diag}(\lambda_1, \dots, \lambda_m)$, $N = (n_{ij})$. Entonces

$$\begin{aligned} \varphi_2(M + N) &= \sum_{i < j} \begin{vmatrix} \lambda_i + n_{ii} & n_{ij} \\ n_{ij} & \lambda_j + n_{jj} \end{vmatrix} \\ &= \varphi_2(M) + \varphi_2(N) + \sum_{i < j} (\lambda_i n_{jj} + \lambda_j n_{ii}), \end{aligned}$$

y por otra parte

$$\begin{aligned} \sum_{i < j} (\lambda_i n_{jj} + \lambda_j n_{ii}) &= \sum_{i \neq j} \lambda_i n_{jj} \\ &= \sum_{i, j} \lambda_i n_{jj} - \sum_i \lambda_i n_{ii} \\ &= \left(\sum_i \lambda_i \right) \left(\sum_j n_{jj} \right) - \left(\sum_i \lambda_i n_{ii} \right) \\ &= \varphi_1(M) \varphi_1(N) - \varphi_1(M N), \end{aligned}$$

con lo que se acaba. \square

Corolario 5.3 *φ_2 es superaditiva en \mathcal{M} , esto es, si M y N son semidefinido positivas se tiene la desigualdad*

$$\varphi_2(M + N) \geq \varphi_2(M) + \varphi_2(N)$$

Demostración: Resulta inmediatamente de la Proposición 5.9 que $\varphi_2(M + N) \geq \varphi_2(M) + \varphi_2(N)$, puesto que la diferencia es

$$\sum_{i < j} (\lambda_i n_{jj} + \lambda_j n_{ii}),$$

y tanto los λ_i como los n_{ij} son mayores o iguales que cero por ser las matrices semidefinido positivas. \square

Proposición 5.10 Si $N \leq M$ entonces φ_2 es convexa en el segmento que las une

Demostración: Sean $M_\alpha = \alpha M + (1 - \alpha)N$ y $M_\beta = \beta M + (1 - \beta)N$, y supongamos $\alpha \geq \beta$. Entonces

$$\begin{aligned} M_\alpha - M_\beta &= [\alpha M + (1 - \alpha)N] - [\beta M + (1 - \beta)N] \\ &= (\alpha - \beta) [M - N] \geq 0. \end{aligned}$$

Además

$$\begin{aligned} &\lambda \varphi_2(M_\alpha) + (1 - \lambda) \varphi_2(M_\beta) - \varphi_2[\lambda M_\alpha + (1 - \lambda)M_\beta] \\ &= \lambda \varphi_2(M_\alpha) + (1 - \lambda) \varphi_2(M_\beta) - \{\lambda^2 \varphi_2(M_\alpha) + (1 - \lambda)^2 \varphi_2(M_\beta) \\ &\quad + \lambda(1 - \lambda)[\varphi_1(M_\alpha)\varphi_1(M_\beta) - \varphi_1(M_\alpha M_\beta)]\} \\ &= \lambda(1 - \lambda)\{\varphi_2(M_\alpha) + \varphi_2(M_\beta) - [\varphi_1(M_\alpha)\varphi_1(M_\beta) - \varphi_1(M_\alpha M_\beta)]\} \\ &= \lambda(1 - \lambda) \varphi_2(M_\alpha - M_\beta), \end{aligned}$$

donde hemos usado el desarrollo (5.10). Por ser $M_\alpha - M_\beta$ semidefinido positiva $\varphi_2(M_\alpha - M_\beta) \geq 0$, con lo que se concluye. \square

Corolario 5.4 Dada cualquier $M \in \mathcal{M}$, φ_2 es convexa en el conjunto $\{N \in \mathcal{M} : M \text{ y } N \text{ son comparables}\}$.

Lema 5.4 Se verifica

$$\varphi_1(M)\varphi_1(N) - \varphi_1(MN) \geq 2\varphi_2(M)^{\frac{1}{2}}\varphi_2(N)^{\frac{1}{2}}$$

Demostración: Sea $M = \text{diag}(\lambda_1, \dots, \lambda_m)$ y $N = (n_{ij})$, simétrica. Por una parte

$$\begin{aligned} [\varphi_1(M)\varphi_1(N) - \varphi_1(MN)]^2 &= \left(\sum_{i \neq k} \lambda_i n_{kk} \right)^2 \\ &= \sum_{i \neq k} \lambda_i^2 n_{kk}^2 + \sum (\lambda_i n_{kk})(\lambda_j n_{ll}) = \mathcal{X} \end{aligned}$$

El segundo sumando es para $\{i \neq k \wedge j \neq l\} \wedge \{i \neq j \vee k \neq l\}$, mientras que

$$\begin{aligned} \varphi_2(M)\varphi_2(N) &= \left(\sum_{i < j} \lambda_i \lambda_j \right) \left(\sum_{k < l} [n_{kk}n_{ll} - n_{kl}^2] \right) \\ &= \underbrace{\left(\sum \lambda_i \lambda_j n_{kk}n_{ll} \right)}_{\mathcal{Y}} - \underbrace{\left(\sum \lambda_i \lambda_j n_{kl}^2 \right)}_{\mathcal{Z}}, \end{aligned}$$

Puesto que $\mathcal{Z} \geq 0$ concluiremos si probamos que $\mathcal{X} \geq 4\mathcal{Y}$. Para ello vamos a ver de qué formas pueden aparecer los sumandos de \mathcal{Y} :

- $\lambda_i \lambda_j n_{kk} n_{ll}$, con $\text{Card}\{i, j, k, l\}=4$ (todos distintos). En este caso, cada uno de los sumandos aparece repetido cuatro veces en \mathcal{X} , ya que este contiene

$$2(\lambda_i n_{kk})(\lambda_j n_{ll}) + 2(\lambda_i n_{ll})(\lambda_j n_{kk})$$

- $\lambda_i \lambda_j n_{ii} n_{ll}$, con $i \neq j, l$. Esta vez el sumando sólo aparece dos veces en \mathcal{X} :

$$2(\lambda_i n_{ll})(\lambda_j n_{ii}).$$

Pero en este caso utilizamos que \mathcal{X} contiene el término $(\lambda_i n_{ll} + \lambda_j n_{ii})^2$, y entonces

$$(\lambda_i n_{ll} + \lambda_j n_{ii})^2 - 4 \lambda_i \lambda_j n_{ii} n_{ll} = (\lambda_i n_{ll} - \lambda_j n_{ii})^2 \geq 0,$$

todo lo cual prueba que $\mathcal{X} \geq 4 \mathcal{Y}$, con lo que se termina. \square

Proposición 5.11 φ_2 es log-cóncava.

Demostración:

$$\begin{aligned} &\log \varphi_2[\alpha M + (1 - \alpha)N] \\ &= \log \left\{ \alpha^2 \varphi_2(M) + (1 - \alpha)^2 \varphi_2(N) + \alpha(1 - \alpha) [\varphi_1(M)\varphi_1(N) - \varphi_1(MN)] \right\} \\ &\geq \log \left[\alpha^2 \varphi_2(M) + (1 - \alpha)^2 \varphi_2(N) + 2 \alpha(1 - \alpha) \varphi_2(M)^{1/2} \varphi_2(N)^{1/2} \right], \end{aligned}$$

donde hemos utilizado la Proposición 5.9 en la igualdad y el Lema 5.4 en la desigualdad. Además la expresión anterior se puede poner como

$$\log \left[\alpha \varphi_2(M)^{1/2} + (1 - \alpha) \varphi_2(N)^{1/2} \right]^2$$

$$\begin{aligned}
&= 2 \log \left[\alpha \varphi_2(M)^{1/2} + (1 - \alpha) \varphi_2(N)^{1/2} \right] \\
&\geq 2 \left[\alpha \log \varphi_2(M)^{1/2} + (1 - \alpha) \log \varphi_2(N)^{1/2} \right] \\
&= \alpha \log \varphi_2(M) + (1 - \alpha) \log \varphi_2(N),
\end{aligned}$$

siendo la desigualdad final debida a la concavidad del logaritmo. \square

Observación 5.5 *Una propiedad deseable en las funciones criterio es la convexidad. Así, tomando como referencia la D -optimización, en la que $\Phi_D(M) = \log \det(M^{-1}) = -\log \det(M)$, podría considerarse como función criterio $\psi(M) = -\log \varphi_2(M)$, que como se acaba de ver es convexa.*

Proposición 5.12 *Si definimos $\psi_2(M) = \varphi_2(M)^{\frac{1}{2}}$, ésta verifica las propiedades*

1. ψ_2 es homogénea:

$$\psi_2(\alpha A) = \alpha \psi_2(M)$$

2. ψ_2 es superaditiva:

$$\psi_2(M + N) \geq \psi_2(M) + \psi_2(N)$$

Demostración: La primera afirmación es inmediata. Veamos la segunda:

$$\begin{aligned}
\psi_2(M + N) &= \varphi_2(M + N)^{\frac{1}{2}} \\
&= [\varphi_2(M) + \varphi_2(N) + \varphi_1(M)\varphi_1(N) - \varphi_1(MN)]^{\frac{1}{2}} \\
&\geq \left[\varphi_2(M) + 2\varphi_2(M)^{\frac{1}{2}}\varphi_2(N)^{\frac{1}{2}} + \varphi_2(N) \right]^{\frac{1}{2}} \\
&= \left[\left(\varphi_2(M)^{\frac{1}{2}} + \varphi_2(N)^{\frac{1}{2}} \right)^2 \right]^{\frac{1}{2}} \\
&= \varphi_2(M)^{\frac{1}{2}} + \varphi_2(N)^{\frac{1}{2}} \\
&= \psi_2(M) + \psi_2(N)
\end{aligned}$$

La segunda igualdad se deriva de la Proposición 5.9, y la desigualdad del Lema 5.4. \square

Observación 5.6 *A partir de la descomposición de $\varphi_2(M + N)$ que se ha visto en la Proposición 5.9 y la fórmula de recurrencia (5.5) resulta sencillo obtener una descomposición similar para las restantes funciones características. A continuación figuran como ejemplo las fórmulas calculadas para φ_3 y φ_4 :*

•

$$\begin{aligned} \varphi_3(M + N) = \varphi_3(M) + \varphi_3(N) + & \begin{array}{l} +\varphi_2(M)\varphi_1(N) \\ +\varphi_2(N)\varphi_1(M) \end{array} \\ & \begin{array}{l} -\varphi_1(M)\varphi_1(MN) \quad +\varphi_1(M^2N) \\ -\varphi_1(N)\varphi_1(MN) \quad +\varphi_1(MN^2) \end{array} \end{aligned}$$

•

$$\begin{aligned} \varphi_4(M + N) = \varphi_4(M) + \varphi_4(N) & \begin{array}{l} +\varphi_3(M)\varphi_1(N) \quad -\varphi_2(M)\varphi_1(MN) \\ +\varphi_3(N)\varphi_1(M) \quad -\varphi_2(N)\varphi_1(MN) \end{array} \\ & \begin{array}{l} +\varphi_1(M)[\varphi_1(M^2N) + \varphi_1(MN^2)] \quad -\varphi_1(M^3N) \\ +\varphi_1(N)[\varphi_1(M^2N) + \varphi_1(MN^2)] \quad -\varphi_1(M^2N^2) \\ \quad \quad \quad \quad \quad \quad \quad -\varphi_1(MN^3) \end{array} \\ & +\varphi_2(M)\varphi_2(N) + \varphi_2(MN) - \varphi_1(M)\varphi_1(N)\varphi_1(MN) \end{aligned}$$

5.4 Interpretación geométrica

Como ya se ha comentado previamente, las longitudes de los ejes del elipsoide de confianza de las estimaciones de los parámetros del modelo son proporcionales a las raíces cuadradas de los valores propios de la matriz de covarianzas (Atkinson y Donev (1992), pgs 42, 48-53). Supongamos que $m = 3$ para poder visualizar nuestros comentarios. El volumen de dicho elipsoide resulta así proporcional al producto de sus ejes, y por tanto minimizar el producto de los valores propios significa minimizar el volumen del elipsoide (D -optimización). De la misma forma, el área exterior de la “caja” que contiene al elipsoide será la suma de los productos de los ejes tomados de dos en dos. Así el criterio Φ_{Ch_2} minimiza un cierto promedio de las áreas de las proyecciones del elipsoide sobre los tres planos canónicos. Además, estas proyecciones pueden verse como las regiones de confianza de las estimaciones

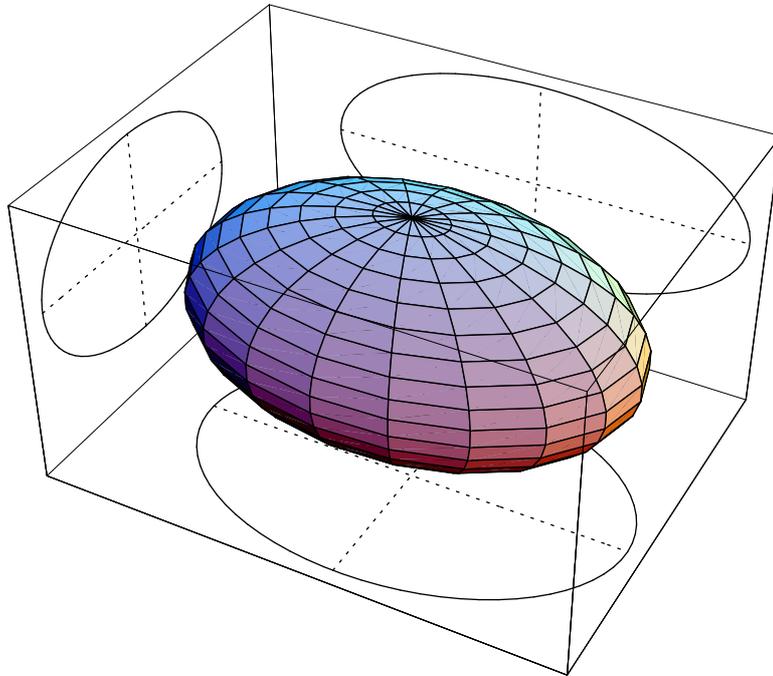


Figura 5.1: Interpretación geométrica del Φ_{Ch_2} -óptimo para el modelo cuadrático

de los distintos pares de parámetros, y el diseño Φ_{Ch_2} -óptimo como el que minimiza el promedio de las áreas de todas estas regiones de confianza. En general, el diseño Φ_{Ch_k} -óptimo minimizaría así el promedio de las regiones de confianza de los estimadores de k parámetros del modelo. Una representación de esta idea para $m = 3$ se presenta en la Figura 5.1.

Capítulo 6

Cálculo aproximado de los Φ_{Ch_k} -óptimos

6.1 Descripción del algoritmo general

Con el fin de simplificar los cálculos, de ahora en adelante tomaremos como funciones criterio las anteriormente definidas $\Phi_{Ch_k}(M) = \varphi_k(M^{-1})$. Se va a utilizar el algoritmo general descrito en el libro “Foundations of Optimum Experimental Design”, Pázman (1986), pg 157. Una vez elegido el modelo y el intervalo del diseño se siguen los siguientes pasos:

1. Se toma como diseño inicial cualquier ξ_0 tal que $\det M(\xi_0) \neq 0$, siendo $M(\xi_0)$ la matriz de información del diseño.
2. En cada etapa se modifica el diseño anterior, de la forma siguiente:

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

donde los β_n tienen que cumplir las condiciones

$$\beta_n \in (0, 1), \quad \sum_{n=0}^{\infty} \beta_n = \infty, \quad \lim_{n \rightarrow \infty} \beta_n = 0$$

y x_n es el punto donde se alcanza el mínimo de la función

$$x \in X \longrightarrow f^t(x) \nabla \Phi[M(\xi_n)]f(x)$$

siendo X el espacio del diseño, $f(x)$ el modelo y $\nabla\Phi[M(\xi)]$ el gradiente de la función criterio Φ aplicado a la matriz de información del diseño ξ .

El algoritmo ha sido implementado en el ordenador usando el programa de cálculo matemático *Mathematica*, y por ejemplo los resultados mostrados en la Tabla 6.1 se han obtenido tras la realización de 10000 iteraciones del algoritmo. El programa consta de una serie de funciones “independientes”, como por ejemplo las que calculan la matriz de información de un diseño dado, el gradiente de una función criterio o el mínimo de una función en un intervalo. Aunque utiliza la función *FindMinimum* del *Mathematica* no es igual que ella, ya que ha de permitir alcanzar los puntos mínimos en los extremos del intervalo. Además incluye un parámetro de precisión. La función principal es la que busca el óptimo. Esta incluye un “refinamiento” cada cierto número de iteraciones, mediante el que se eliminan los puntos cuyo peso ha llegado a ser despreciable en ese momento. Mediante un parámetro de *proximidad* se añade entonces su peso al punto más cercano, produciendo otro punto que es la media de ambos, o bien se reparte entre todos si no hay ninguno suficientemente próximo. Todos estos parámetros: el número de iteraciones en las que se produce el refinamiento, el límite a partir del cual un punto pasa a ser considerado de peso despreciable, la proximidad entre puntos, así como algunos otros pueden ser fijados por el usuario. El Apéndice A recoge una descripción más detallada del programa.

También se tiene la posibilidad de fijar un *parámetro de cercanía* que indicaría que nuestro diseño está “suficientemente cerca” del óptimo, y que por tanto ya no haría falta seguir ejecutando el algoritmo. Esto es, establecer una condición de parada para el algoritmo, como se ha visto en (3.3).

6.2 Modelo polinómico

El modelo polinómico es seguramente el más usado en todas las disciplinas, debido a su sencillez y a que se convierte en una aproximación fiable de funciones mucho más complicadas o inmanejables en la práctica. En este trabajo se han calculado los Φ_{Ch_k} -óptimos para los modelos polinómicos de grados mas pequeños (de dos a seis), y los resultados están representados en la Tabla 6.1. Aunque los óptimos para el primer y último criterio de cada modelo son

conocidos y están reflejados en la literatura, se ha preferido incluir también en estos casos los resultados obtenidos con el ordenador, por ser casos particulares de un cálculo más general. En particular es sabido (ver Pázman 1986, pg 179) que el D -óptimo se puede obtener explícitamente para el modelo polinómico, ya que tiene soporte en las raíces de los polinomios de Legendre. Esto, unido a otros resultados (Pázman 1986, pg 177) que demuestran que cuando el soporte del D -óptimo tiene un número de puntos igual al número de parámetros todos esos puntos tienen el mismo peso, concluye el cálculo del óptimo para este criterio.

Lo primero que se observa es que los Φ_{Ch_k} -óptimos tienen soporte en tantos puntos como parámetros tiene el modelo, resultado ya conocido para $k = 1$ y $k = m$ (A - y D -optimización respectivamente). Además es fácil apreciar que los primeros Φ_{Ch_k} -óptimos dan menos importancia a los puntos extremos del intervalo, importancia que va aumentando con k hasta llegar al Φ_{Ch_m} -óptimo (es decir, el D -óptimo), para el que los m puntos escogidos son igualmente importantes. Por lo demás, parece existir una transición “suave” entre el A -óptimo y el D -óptimo a través de los criterios intermedios.

6.2.1 Eficiencias

La Tabla 6.2 representa las eficiencias de los diseños calculados anteriormente respecto de las propias funciones criterio características. En la casilla (i, j) de la parte central de la tabla figura pues la eficiencia del diseño Φ_{Ch_i} -óptimo respecto de la función criterio Φ_{Ch_j} , para los distintos grados.

Como era de esperar, para cada modelo la eficiencia del Φ_{Ch_k} -óptimo va disminuyendo al aumentar k , pero es fácil observar que para valores pequeños de k esta disminución es mucho más acusada que en los casos de k cercano a m . Cuando calculamos la eficiencia media, o la media cuadrática de las eficiencias de cada óptimo, sea para todos los criterios característicos o sólo para los conocidos de A - y D -optimización, encontramos que las cifras mayores se alcanzan para valores grandes de k . Pero es sorprendente observar que el mejor valor no corresponde a $k = m$, sino, dependiendo de los casos, a $k = m - 1$ ó $k = m - 2$. Esto significa que si consideramos por ejemplo el modelo cuadrático, y estamos interesados en minimizar el promedio de las varianzas de los estimadores sin tener en cuenta las covarianzas, deberíamos

n	k	Φ_{Ch_k} -óptimos para el modelo polin. de grado n						
2	1	-1.	0	1.				
	2	0.25	0.5	0.25				
		-1.	0	1.				
	3	0.297	0.407	0.297				
3	1	-1.	-0.464	0.464	1.			
		0.15	0.35	0.35	0.15			
	2	-1.	-0.424	0.424	1.			
		0.173	0.327	0.327	0.173			
	3	-1.	-0.435	0.435	1.			
		0.215	0.285	0.285	0.215			
	4	-1.	-0.447	0.447	1.			
		0.25	0.25	0.25	0.25			
4	1	-1.	-0.677	0	0.677	1.		
		0.105	0.25	0.29	0.25	0.105		
	2	-1.	-0.643	0	0.643	1.		
		0.116	0.256	0.256	0.256	0.116		
	3	-1.	-0.633	0	0.633	1.		
		0.139	0.232	0.257	0.232	0.139		
	4	-1.	-0.643	0	0.643	1.		
		0.17	0.216	0.228	0.216	0.17		
	5	-1.	-0.655	0	0.655	1.		
		0.20	0.20	0.20	0.20	0.20		
5	1	-1.	-0.789	-0.291	0.291	0.789	1.	
		0.08	0.187	0.232	0.232	0.187	0.08	
	2	-1.	-0.767	-0.267	0.267	0.767	1.	
		0.086	0.198	0.216	0.216	0.198	0.086	
	3	-1.	-0.749	-0.28	0.28	0.749	1.	
		0.099	0.195	0.206	0.206	0.195	0.099	
	4	-1.	-0.747	-0.278	0.278	0.747	1.	
		0.116	0.179	0.205	0.205	0.179	0.116	
	5	-1.	-0.756	-0.281	0.281	0.756	1.	
		0.141	0.173	0.186	0.186	0.173	0.141	
	6	-1.	-0.765	-0.285	0.285	0.765	1.	
		0.167	0.167	0.167	0.167	0.167	0.167	
6	1	-1.	-0.853	-0.479	0	0.479	0.853	1.
		0.065	0.148	0.185	0.205	0.185	0.148	0.065
	2	-1.	-0.839	-0.451	0	0.451	0.839	1.
		0.069	0.156	0.182	0.186	0.182	0.156	0.069
	3	-1.	-0.824	-0.455	0	0.455	0.824	1.
		0.076	0.162	0.166	0.191	0.166	0.162	0.076
	4	-1.	-0.815	-0.461	0	0.461	0.815	1.
		0.086	0.155	0.171	0.175	0.171	0.155	0.086
	5	-1.	-0.817	-0.46	0	0.46	0.817	1.
		0.1	0.145	0.168	0.173	0.168	0.145	0.1
	6	-1.	-0.823	-0.464	0	0.464	0.823	1.
		0.121	0.144	0.156	0.158	0.156	0.144	0.121
	7	-1.	-0.83	-0.469	0	0.469	0.83	1.
		0.143	0.143	0.143	0.143	0.143	0.143	0.143

Tabla 6.1: Óptimos característicos para los primeros modelos polinómicos

		Eficiencias						
n	k	1	2	3	4	5	6	7
2	1		98.5	94.5				
	2	98.8		98.9				
	3	96.2	98.9					
3	1		98.9	96.1	91.7			
	2	99.2		98.5	95.0			
	3	97.9	98.8		99.0			
	4	96.1	96.0	99.0				
4	1		99.0	97.4	94.6	90.7		
	2	99.3		99.2	96.9	93.2		
	3	98.5	99.4		99.0	96.3		
	4	97.7	98.0	99.1		99.1		
	5	96.6	96.0	96.8	99.2			
5	1		99.1	97.7	96.4	94.0	90.6	
	2	99.4		99.3	98.2	95.9	92.5	
	3	98.7	99.4		99.6	97.9	94.9	
	4	98.2	98.8	99.6		99.2	97.0	
	5	97.7	97.5	98.5	99.3		99.3	
	6	97.0	96.3	96.6	97.3	99.3		
6	1		99.2	98.0	96.8	95.8	93.8	90.7
	2	99.4		99.4	98.4	97.4	95.4	92.3
	3	98.8	99.5		99.6	98.8	97.1	94.2
	4	98.5	98.9	99.6		99.7	98.3	95.8
	5	98.2	98.4	99.1	99.7		99.4	97.6
	6	97.8	97.7	98.2	98.7	99.4		99.4
	7	97.3	96.7	96.7	97.0	97.8	99.4	

Tabla 6.2: Eficiencias (%) de los Φ_{Ch_k} -óptimos para los modelos polinómicos de grado n

elegir el A -óptimo ($k = 1$). Si nuestro interés fuera minimizar el volumen del elipsoide de confianza de las estimaciones de los parámetros elegiríamos el diseño D -óptimo ($k = 3$). Pero si nuestras intenciones no están definidas muy claramente, la mejor opción debería ser elegir el diseño Φ_{Ch_2} -óptimo, que es bastante eficiente para ambos casos. Análogamente sucede con los modelos de otros grados. Evidentemente esto podría hacerse también con criterios compuestos. En el caso en consideración consistiría en buscar una combinación convexa adecuada de los criterios de A y D -optimización. Como es bien sabido, su tratamiento no es numéricamente sencillo, de modo que los criterios característicos intermedios ofrecen una solución rápida y eficiente.

6.2.2 Relaciones

Los Φ_{Ch_k} -óptimos que hemos hallado para el modelo polinómico están fuertemente relacionados entre sí. Esto sucede cuando prestamos atención tanto al soporte como a los pesos, cuando fijamos el grado en el modelo y hacemos variar k , o cuando dejamos fijo k y examinamos los Φ_{Ch_k} -óptimos para distintos grados. A continuación comentaremos algunos gráficos que ilustran estos conceptos. En todos los óptimos que aparecen, cada punto del soporte correspondiente está representado con un grosor proporcional a su peso.

En la Figura 6.1 están representados todos los óptimos característicos para modelos de dos grados distintos. Queda patente la conexión entre ellos y la suave transición entre los Φ_{Ch_k} -óptimos extremos ($k = 1$ y $k = m$) a través de los intermedios.

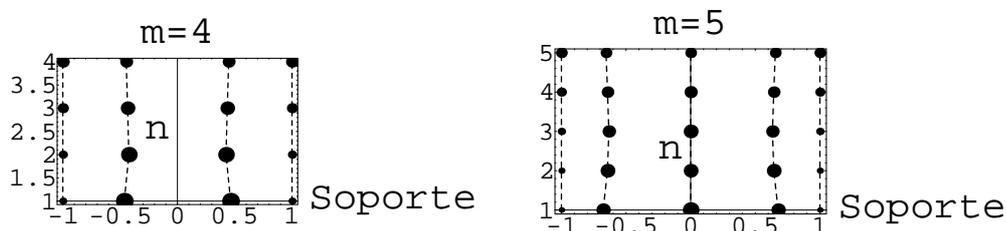


Figura 6.1: Φ_{Ch_k} -óptimos para los modelos de grados 3 y 4

Obsérvese la Figura 6.2, en la que hemos fijado k y dejamos variar m . Los Φ_{Ch_k} -óptimos, representados en rectas paralelas al eje de abscisas, dibujan curvas bien definidas. En ellas el grosor de los puntos, que representa su peso, también parece disminuir de una manera regular y suave.

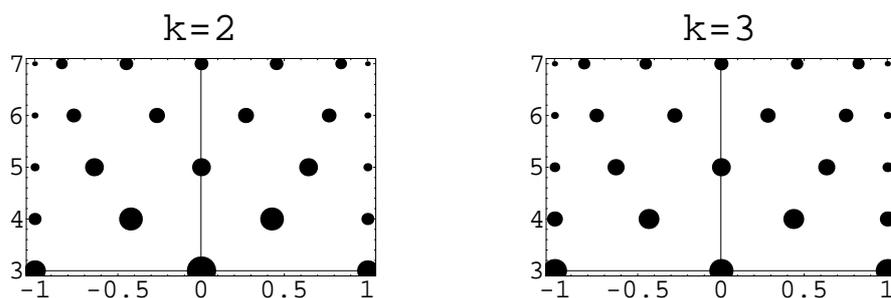


Figura 6.2: Φ_{Ch_2} y Φ_{Ch_3} -óptimos para modelos de distintos grados

Los pesos no escapan a la relación que impregna todo el conjunto, lo que se pone de manifiesto en la Figura 6.3, en la que vemos como disminuyen de manera regular los pesos de los puntos del soporte cuando tomamos k fijo y observamos distintos valores de n . Ordenando los puntos del soporte, el primero corresponde a la curva inferior, el segundo a la siguiente, y así sucesivamente.

6.2.3 Cálculo aproximado de los Φ_{Ch_k} -óptimos

A continuación presentamos un método para el cálculo aproximado de diseños óptimos característicos para el modelo polinómico, basado en una de las relaciones más sobresalientes representadas anteriormente, la que se observa cuando fijamos k y representamos los Φ_{Ch_k} -óptimos característicos para distintos n . Los resultados que se expondrán a continuación pueden consultarse también en Rodríguez Díaz y López Fidalgo (2000).

En particular nos hemos centrado en el caso $k = 1$ (A -optimización), y para éste nuestras afirmaciones pueden apreciarse mejor en la Figura 6.4, que incluye los Φ_{Ch_1} -óptimos para los modelos polinómicos de grados entre 2 y 12, cada uno representado en la ordenada n , donde n es el grado del

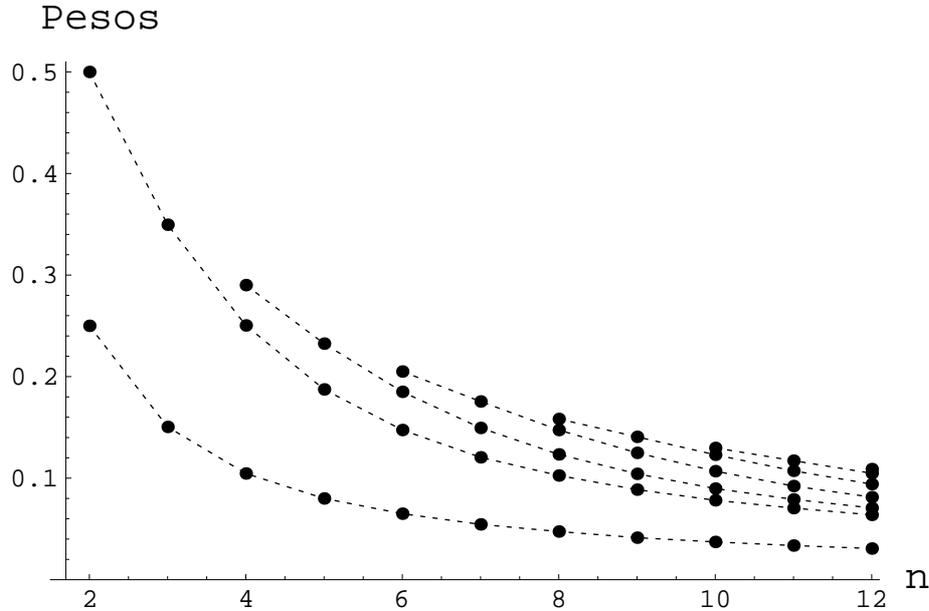


Figura 6.3: Disminución de los pesos de los puntos del soporte del A -óptimo ($k = 1$) al aumentar el grado

modelo. Los datos numéricos correspondientes a estos diseños se encuentran en la Tabla 6.3.

Fijándonos en este patrón que parece regir los Φ_{Ch_1} -óptimos, parece lógico intentar extrapolar los puntos del soporte del Φ_{Ch_1} -óptimo para los modelos polinómicos de grados superiores, utilizando todas las curvas que van definiendo los correspondientes a los grados conocidos. Así pues hemos intentado calcular estas curvas, y en principio se han utilizado los modelos

$$x_{1_i} = \pm \left[1 - \left(\frac{2i}{n} \right)^{c_i} \right], \quad x_{2_i} = \pm \left[1 - \left(\frac{2i}{n} \right)^{d_i} \right], \quad i = 0, 1, 2, \dots, \left[\frac{n}{2} \right],$$

que son bastante sencillos, y cuyo parámetro se puede aproximar razonablemente por una función lineal de i . Es claro a partir del dibujo que las secuencias $\{c_i\}$ y $\{d_i\}$ tienen que ser decrecientes. Se utilizó primeramente regresión no lineal para estimar c_i y d_i a partir de los puntos de los soportes correspondientes hasta grado 12. Una vez calculados los diferentes estimadores de c_i y d_i , $i = 1, \dots, 5$, se efectuó regresión lineal para relacionar

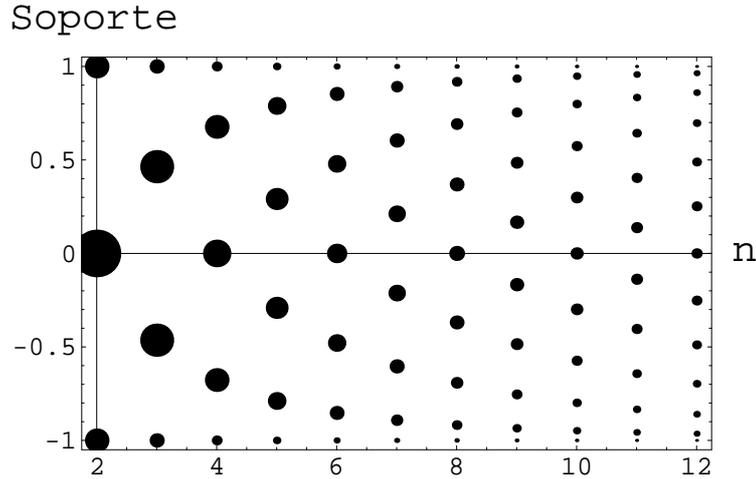


Figura 6.4: A -óptimos para los modelos polinómicos de grados 2 a 12

ambos con el parámetro i , obteniendo las fórmulas $c_i = 1.741 - 0.0297 i$ y $d_i = 0.325 - 0.03 i$, que serán los Modelos I y II respectivamente. La representación gráfica de estas curvas muestra que la primera se desvía ligeramente por debajo de los verdaderos valores, mientras que con la segunda sucede lo contrario. Esto nos sugirió que una combinación convexa $x_{3_i} = (1 - \alpha)x_{1_i} + \alpha x_{2_i}$, $0 \leq \alpha \leq 1$ de ambos modelos podría proporcionar buenos resultado (Modelo III).

Con este sistema podemos aspirar a obtener los puntos soporte del A -óptimo correspondiente para grados altos del polinomio. A pesar de que la representación de los pesos (grosor de los puntos) invita a intentar repetir el mismo argumento con ellos, este extremo no es necesario, ya que si tenemos en cuenta que el soporte de todos los Φ_{Ch_k} -óptimos (y en particular los A -óptimos) para los modelos polinómicos tiene un número de puntos igual al número de parámetros, podemos utilizar el método descrito por Pukelsheim y Torsney (1991) para hallar los pesos óptimos que corresponden a un soporte dado. Por tanto parece lógico intentar obtener el soporte de los nuevos A -óptimos para el modelo polinómico en base a la relación (representada por las curvas) implícita entre los A -óptimos para distintos grados del modelo, y luego calcular los pesos óptimos utilizando el proceso descrito en el artículo mencionado. En este último se calculan además los mejores pesos desde el punto de vista de A -optimización para el soporte dado por la distribución

n	A-óptimos para el modelo polinómico de grado n												
2	-1	0	1										
	.25	.5	.25										
3	-1	-.464	.464	1									
	.15	.35	.35	.15									
4	-1	-.677	0	.677	1								
	.105	.25	.29	.25	.105								
5	-1	-.789	-.291	.291	.789	1							
	.08	.187	.232	.232	.187	.0801							
6	-1	-.853	-.479	0	.479	.853	1						
	.065	.147	.185	.205	.185	.148	.065						
7	-1	-.892	-.604	-.212	.212	.604	.892	1					
	.0545	.121	.149	.176	.176	.149	.121	.0545					
8	-1	-.918	-.692	-.369	0	.369	.692	.918	1				
	.0475	.103	.123	.147	.158	.147	.123	.103	.0475				
9	-1	-.935	-.754	-.485	-.167	.167	.485	.754	.935	1			
	.041	.089	.104	.125	.141	.141	.125	.104	.089	.041			
10	-1	-.948	-.799	-.574	-.299	0	.299	.574	.799	.948	1		
	.037	.078	.09	.107	.123	.13	.123	.107	.09	.078	.037		
11	-1	-.957	-.834	-.643	-.404	-.138	.138	.404	.643	.834	.957	1	
	.034	.071	.079	.092	.107	.117	.117	.107	.092	.079	.071	.034	
12	-1	-.964	-.86	-.697	-.489	-.252	0	.252	.489	.697	.86	.964	1
	.031	.064	.071	.081	.094	.105	.109	.105	.094	.081	.071	.061	.031

Tabla 6.3: A-óptimos para los primeros modelos polinómicos

arcoseno. Ésta, aunque introducida por Fedorov (1972) para D -optimización, es utilizada con muy buenos resultados por Pukelsheim y Torsney para A -optimización, obteniendo unas eficiencias bastante buenas, incluso mejores que las D -eficiencias correspondientes al mismo soporte con los mejores pesos para D -optimización.

Las eficiencias para los tres modelos de puntos del soporte, junto con sus pesos óptimos, se muestran en la Tabla 6.4. Para el Modelo III se ha tomado $\alpha = 0.5$, y claramente proporciona unas eficiencias muy destacables para los primeros modelos polinómicos. Con el fin de comparar eficiencias, se ha añadido el soporte del arcoseno junto con sus pesos óptimos (Modelo IV).

En principio no podríamos usar el mismo argumento para D -optimización. La razón es que mientras el A -óptimo coincide siempre con el Φ_{Ch_1} -óptimo, el D -óptimo se corresponde con el Φ_{Ch_m} -óptimo, donde $m = n + 1$ depende del grado del modelo polinómico. Sin embargo, haciendo la misma repre-

n	3	4	5	6	7	8	9	10	11	12
I	98.9	99.5	99.4	99.5	99.2	98.9	98.3	97.5	96.4	95.0
II	98.9	98.8	99.3	99.8	99.9	99.5	98.4	96.5	93.8	90.3
III	100.	100.	99.7	99.7	99.7	99.8	99.8	99.9	100.	99.9
IV	98.9	98.6	98.6	98.5	98.5	98.5	98.5	98.5	98.5	98.5

Tabla 6.4: Eficiencias (%) para A -optimización

sentación para $k = m$ obtenemos unas curvas similares (en realidad sucede lo mismo para $k = m - 1, k = m - 2, \dots$), y se han utilizado los mismos Modelos para aproximarlas. Esta vez hemos usado los puntos soporte de los D -óptimos hasta grado 40, e $i = 1, 2, \dots, 11$. Los primeros diseños están representados en la Tabla 6.5. Los valores de las estimaciones de los parámetros fueron en esta ocasión $c_i = 1.7$ y $d_i = 0.229 - 0.0086 i$, mientras que todos los pesos se han tomado como $1/m$. Las eficiencias correspondientes se muestran en la Tabla 6.6.

Observación 6.1 *Hemos estudiado una aproximación a los puntos del soporte utilizando regresión. Se podría intentar un acercamiento similar para los pesos óptimos, cuando el proceso descrito por Pukelsheim y Torsney (1991) fuera demasiado complicado, o incluso que no se pudiera aplicar, como ocurriría en los casos en que el número de puntos del soporte que manejásemos fuera mayor que el número de parámetros del modelo.*

6.3 Otros modelos

Aunque en este trabajo nos hemos centrado en el modelo polinómico –por razones que ya se han explicado–, el uso de los criterios característicos que se han presentado no se restringe a este ámbito. Querríamos por tanto antes de finalizar fijarnos en modelos distintos a los que aplicaremos las nuevas funciones criterio.

6.3.1 Modelo exponencial

Muy utilizados son también los modelos representados como sumas de exponenciales, pero lineales en los parámetros. Vamos a calcular los diseños

n	D -óptimos para el modelo polinómico de grado n												
2	-1	0	1										
	.333	.333	.333										
3	-1	-.447	.447	1									
	.25	.25	.25	.25									
4	-1	-.655	0	.655	1								
	.2	.2	.2	.2	.2								
5	-1	-.765	-.285	.285	.765	1							
	.167	.167	.167	.167	.167	.167							
6	-1	-.83	-.469	0	.469	.83	1						
	.143	.143	.143	.143	.143	.143	.143						
7	-1	-.872	-.592	-.209	.209	.592	.872	1					
	.125	.125	.125	.125	.125	.125	.125	.125					
8	-1	-.9	-.677	-.363	0	.363	.677	.9	1				
	.111	.111	.111	.111	.111	.111	.111	.111	.111				
9	-1	-.92	-.739	-.478	-.165	.165	.478	.739	.92	1			
	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1			
10	-1	-.934	-.784	-.565	-.296	0	.296	.565	.784	.934	1		
	.091	.091	.091	.091	.091	.091	.091	.091	.091	.091	.091		
11	-1	-.945	-.819	-.633	-.4	-.137	.137	.4	.633	.819	.945	1	
	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	
12	-1	-.953	-.846	-.686	-.483	-.249	0	.249	.483	.686	.846	.953	1
	.077	.077	.077	.077	.077	.077	.077	.077	.077	.077	.077	.077	.077

Tabla 6.5: D -óptimos para los primeros modelos polinómicos

óptimos característicos del modelo

$$\eta(x, \theta) = \theta_1 + \theta_2 e^x + \theta_3 e^{-x} \tag{6.1}$$

en el intervalo $[0,1]$. Los resultados se muestran en la Tabla 6.7.

Las eficiencias de estos óptimos respecto de las propias funciones criterio características están representadas en la Tabla 6.8. Para este modelo, el Φ_{Ch_3} -óptimo parece el más eficiente globalmente aunque, como puede observarse, los valores correspondientes al Φ_{Ch_2} -óptimo se acercan mucho a éstos.

6.3.2 Modelo compartimental

El *modelo compartimental* es un modelo no lineal frecuentemente usado en biocinética y farmacocinética. Su elección se debe a varias razones:

n	3	4	5	6	7	8	9	10	15	20
I	99.2	99.3	99.2	99.4	99.5	99.6	99.6	99.6	99.5	99.0
II	99.4	99.1	99.0	99.2	99.3	99.5	99.7	99.9	99.6	98.0
III	100.	100.	99.9	99.9	99.9	99.9	99.9	99.9	99.8	99.8
IV	99.1	98.6	98.3	98.1	98.0	97.9	97.9	97.9	98.0	98.2

Tabla 6.6: Eficiencias (%) para D -optimización

k	Φ_{Ch_k} -óptimos		
1	0	0.498	1
	0.263	0.516	0.223
2	0	0.482	1
	0.382	0.354	0.266
3	0	0.498	1
	0.331	0.332	0.332

Tabla 6.7: Diseños óptimos característicos para el modelo exponencial

- nuestro deseo de aplicar los aspectos desarrollados a lo largo de los capítulos anteriores a situaciones reales, siendo este uno de los modelos más utilizados en la modelización de multitud de fenómenos,
- su popularidad en los campos ya mencionados, de la que se deriva la abundante literatura existente sobre el tema. A este respecto puede consultarse por ejemplo Atkinson y Donev (1992) pgs 193-195 y 200-207, o Sánchez León (1998), y
- el hecho de ser un modelo no lineal, apenas tratados anteriormente en este trabajo

Se podría dar la siguiente definición, según puede encontrarse en Sanchez León (1998):

Definición 6.1 *Un modelo compartimental es la representación de un sistema físico o biológico que se descompone en un número finito de componentes llamados compartimentos, entre los que se produce una transferencia*

k	Eficiencias	
1	89.30	86.84
2	92.12	98.08
3	93.27	98.75

Tabla 6.8: Eficiencias para el modelo exponencial

de materia. Se dice que es abierto cuando existe intercambio de materia entre algunos compartimentos y el exterior, siendo en otro caso cerrado. Cada uno de los compartimentos se supone homogéneo, y representa una situación de la materia dentro del modelo

El modelo general es una combinación de exponenciales, y se utilizan diversas variantes. Aquí vamos a usar la que aparece en Atkinson y otros (1992); es decir, de ahora en adelante nuestro modelo será

$$\eta(x, \theta) = \theta_3 [e^{-\theta_2 x} - e^{\theta_1 x}]; \quad \theta_2 > \theta_1 > 0, \quad \theta_3 > 0 \quad (6.2)$$

El propio artículo examina un ejemplo que se refiere a la concentración, y , de una determinada sustancia en la sangre en distintos periodos de tiempo x . La estimación inicial de los parámetros se realiza en base a un experimento previo en el que se ha observado y para 18 valores de x , repartidos de una manera creciente entre 10 segundos y 48 minutos. Con estos datos se obtienen por mínimos cuadrados las estimaciones

$$\theta_1^0 = 0.05884, \quad \theta_2^0 = 4.298, \quad \theta_3^0 = 21.80,$$

expresadas con cuatro cifras significativas. Estos serán los valores iniciales que usaremos a la hora de linealizar el modelo. Esto es, tomaremos como modelo

$$f(x) = \left(\frac{\partial \eta(x, \theta)}{\partial \theta_1}, \frac{\partial \eta(x, \theta)}{\partial \theta_2}, \frac{\partial \eta(x, \theta)}{\partial \theta_3} \right) \Bigg|_{\theta = \theta^0},$$

que en nuestro ejemplo resulta

$$f(x) = (21.8 x e^{-0.05884 x}, -21.8 x e^{-4.298 x}, e^{-4.298 x} - e^{-0.05884 x}).$$

Así, convenientemente linealizado, podemos emplear el algoritmo habitual, y utilizar el programa descrito en el Apéndice A para obtener el diseño óptimo.

El diseño así calculado sería un óptimo local, es decir, óptimo en un entorno del punto θ^0 . Los resultados obtenidos para el cálculo de los tres óptimos característicos posibles para el modelo (6.2), tras realizar 10000 iteraciones con el algoritmo para x entre 0 y 50 son los de la Tabla 6.9.

k	Φ_{Ch_k} -óptimos		
1	0.185	1.264	23.426
	0.303	0.581	0.115
2	0.225	1.204	23.100
	0.371	0.527	0.102
3	0.229	1.387	18.405
	0.331	0.336	0.333

Tabla 6.9: Diseños óptimos característicos para el modelo compartimental

Sus eficiencias aparecen representadas en la Tabla 6.10, donde en la casilla (i, j) figura la eficiencia del Φ_{Ch_i} -óptimo respecto de la función criterio Φ_{Ch_j} . En este caso también puede observarse que el Φ_{Ch_2} -óptimo ofrece una alta eficiencia (más del 80%) para los criterios de A y D -optimización, siendo especialmente notable la que alcanza para el primero de ellos, el 93%.

k	Eficiencias	
1	95.59	71.39
2	93.29	80.12
3	78.60	79.07

Tabla 6.10: Eficiencias para el modelo compartimental

Conclusiones

Muchos de los resultados que figuran en el presente trabajo han sido presentados en diversos congresos nacionales e internacionales, y se encuentran publicados en los *proceedings* de estas reuniones. Los dos más notables son sin embargo López Fidalgo y Rodríguez Díaz (1998) y Rodríguez Díaz y López Fidalgo (2000), los cuales han tenido que superar para su publicación la evaluación correspondiente mediante *referees*. En estos dos artículos se encuentran parte de los resultados de los Capítulos 5 y 6, los cuales incorporan conceptos, propiedades y resultados novedosos.

Tras los dos primeros capítulos, en los que se muestran los resultados básicos de la teoría del Diseño Óptimo de Experimentos, se presentan una serie de aportaciones originales, cuyos aspectos más notables son los siguientes:

1. Desarrollo de algoritmos de búsqueda del diseño óptimo sobre conjuntos discretos de puntos. La utilidad de estos algoritmos es más destacable cuando el conjunto de puntos candidatos no es muy grande, ya que el número de estos aumenta el tiempo de cálculo. Son aplicables también en el caso de soportes continuos cuando no se requiere una gran precisión (en términos de número de decimales) en los resultados.
2. Estudio del orden de las matrices de información de los diseños unipuntuales y sus combinaciones convexas, con el objetivo de realizar una buena elección de los diseños iniciales a emplear en los algoritmos mencionados. En general un diseño unipuntual nunca será mejor (en el sentido del Capítulo 4) que uno con más puntos, como puede observarse en el ejemplo final. Sin embargo, si nuestro objetivo es buscar el diseño óptimo utilizando combinaciones convexas de diseños, es claro que los unipuntuales forman la base en la que se ha de empezar. Su elección

resulta problemática, y en muchas ocasiones no es posible establecer una comparación entre ellos. Algunos de los resultados expuestos (por ejemplo la *Caracterización geométrica de su orden parcial*) pueden resultar de utilidad para este propósito.

3. Introducción de una nueva clase de funciones criterio, los llamados Criterios Característicos. Están basados en los coeficientes del polinomio característico de la matriz de información, por lo que resultan una generalización de los conocidos A - y D -optimización. Esta nueva colección de funciones criterio nos lleva a la definición de un criterio más general que engloba tanto a éstos como a la conocida clase de criterios Φ_p . Se obtiene así una clase bidimensional de criterios que generalizan buena parte de los utilizados en la práctica.
4. Relación de numerosas propiedades relativas a los nuevos criterios. Primeramente se estudian las funciones simétricas en los valores propios y se prueba que verifican la propiedad circular, el crecimiento respecto al orden de Loewner, la generación de cada una a partir de las de subíndice menor, la relación entre ellas y las funciones características, etc. Lógicamente a veces es necesario ayudarse de otros resultados menores, en ocasiones de gran complejidad de cálculo. Todos estos resultados se usarán más tarde a la hora de establecer las principales propiedades de las funciones criterio creadas. Entre ellas hay que destacar la demostración de su convexidad y diferenciabilidad, con el cálculo explícito de su gradiente. Para ello se calcula también el de la traza de las potencias de la matriz.
5. Las características de las nuevas funciones criterio hacen que los algoritmos tradicionales sean fácilmente aplicables para la búsqueda de los óptimos correspondientes.
6. Demostración de fórmulas que relacionan los criterios entre sí y evitan el cálculo secuencial de la inversa en cada paso, lo que facilita enormemente la labor computacional.
7. Cálculo de diseños óptimos característicos para modelos polinómicos de diferentes grados. Se comprueba que resultan altamente eficientes al ser comparados entre sí y con los criterios extremos, Φ_A y Φ_D .

8. Obtención de los óptimos característicos para otros modelos, lineales y no lineales, como el exponencial o el compartimental, muy utilizado en algunas ciencias experimentales.
9. Presentación de un método novedoso para el cálculo aproximado de diseños óptimos característicos para los modelos polinómicos. Su eficiencia es evidente al ser comparados con otros óptimos aproximados, y con los conocidos de soporte arco seno.
10. Realización de un programa en *Mathematica* que permite el cálculo de diseños óptimos característicos para modelos lineales.

Apéndice A

Programa CharCriteria

Desarrollamos a continuación una breve descripción del programa **CharCriteria**, implementado con la herramienta de cálculo matemático *Mathematica* para la búsqueda de óptimos característicos. Como ya se ha mencionado, utiliza el algoritmo descrito en el Capítulo 6. Las primeras versiones que se crearon eran muy generales, y el usuario debía suministrar una buena cantidad de parámetros de entrada, incluyendo el modelo (forzosamente lineal), el óptimo característico que se pretendía encontrar, el intervalo de búsqueda y otros parámetros menores de precisión, de parada, etc. Estos primeros programas dieron paso a otros ya más especializados, en los que se fijaron la mayoría de los parámetros iniciales. En algunas versiones se realizaba un número determinado de iteraciones mientras que en otras se empleaban condiciones de parada como las descritas en el Capítulo 3. También variaron las representaciones gráficas realizadas, presentándose en ocasiones información intermedia del desarrollo de los cálculos mientras que otras veces sólo se reflejaba la correspondiente al diseño final obtenido. Por todo ello parece que el procedimiento más adecuado es la descripción del programa general y de las funciones que componen el mismo. En lo que sigue nos referiremos a “parámetros” en sentido informático: datos que utiliza una función para realizar su cometido, y que en ciertas ocasiones son suministrados al realizar la llamada a la función. Las principales funciones implementadas son las siguientes:

- **Matriz de información.** Utiliza como parámetros el diseño, ξ , y el modelo $f(x)$. Procedimiento:
 - Definición, a partir del modelo, de la función $x \mapsto f(x)$. Es nece-

sario hacerlo porque no reconoce el parámetro como función de x , hace falta utilizar otra variable.

- Construcción de la matriz de información del diseño pasado como parámetro, de la manera

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

- **Funciones simétricas**, $\varphi_k(M)$, que reciben k y M como parámetros. Existen distintas maneras de expresarlas, siendo quizá la más clara

$$\varphi_k(M) = \sum_{i_1 < \dots < i_k} \lambda_{i_1} \dots \lambda_{i_k} \quad 1 \leq k \leq m$$

(véase (5.1)), que es la más sencilla desde el punto de vista intuitivo e incluso de implementación con el ordenador, ya que el programa utilizado tiene definidas funciones que calculan los valores e incluso los vectores propios. Todo se hace más complicado sin embargo cuando la matriz que nos ocupa no está completamente definida, sino que depende del valor de alguna variable. Esto ocurre en ocasiones en el proceso algorítmico si por ejemplo queremos incorporar en cada paso un punto adecuado con un peso que buscamos óptimo, del que dependerá la matriz. En este caso resulta más cómodo calcular los coeficientes característicos como

$$\varphi_k(M) = \sum_{i_1 < \dots < i_k} \det[M]_{i_1, \dots, i_k} \quad 1 \leq k \leq m,$$

donde $[M]_{i_1, \dots, i_k}$ denota la matriz construida tomando solamente los elementos de M situados en el cruce de las filas i_1, \dots, i_k con las columnas i_1, \dots, i_k (ver (5.2)). Aquí se ha implementado de esta última manera. Procedimiento:

- Definir $\varphi_0(M) = 1$
- Creación de un bucle (**Do** anidado) que recorra todas las combinaciones posibles de k índices de entre los m . Esta es una tarea complicada de efectuar cuando el número de bucles es decidido en tiempo de ejecución. Para ello se creará una función recursiva, *Rec*, que va añadiendo índices hasta que ya se tienen k , momento en que se construye la matriz $[M]_{i_1, \dots, i_k}$, se toma su determinante y se suma al parámetro *Tot* (total), que es lo que al final se devuelve. Debido a su complejidad presentamos el código correspondiente:

```

Fi[k_,M_] :=
(
  m = Dimensions[M][[1]];
  i[0] = 0;
  Rec[j_] :=
    Do[
      If[j == k,
        Tot += Det[
          M[
            Table[i[1], {1, 1, k}],
            Table[i[1], {1, 1, k}]
          ]
        ],
      Rec[j + 1]
    ],
  {i[j], i[j - 1] + 1, m - k + j}];
  Tot = 0;
  Rec[1];
  Tot
);

```

- **Gradiente de los criterios característicos**, que será conveniente calcular como

$$\nabla \Phi_{Ch_k} = \sum_{i=k}^m \varphi_i(M^{-1}) (-M)^{i-k-1}$$

o alternativamente como

$$\nabla \Phi_{Ch_k} = - \sum_{i=0}^{k-1} \varphi_i(M^{-1}) (-M)^{i-k-1}$$

dependiendo de que k sea o no mayor que $m/2$. El procedimiento es inmediato a partir de las fórmulas y la definición de las funciones φ_k y no será detallado.

- **Beta** calcula el peso óptimo que debe tomar el punto que incorporamos en cada paso del algoritmo. Recibe como parámetros el modelo, el diseño n -ésimo, k y x_n , el punto que hay que incorporar al diseño. Procedimiento:

- Construcción del diseño $(n + 1)$ -ésimo:

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

- Cálculo de la matriz de información correspondiente a ξ_{n+1} , que por supuesto dependerá de β .
- Hallar

$$\Phi_{Ch_k} [M(\xi_{n+1})] = \varphi_k [M^{-1}(\xi_{n+1})].$$

Al ser una matriz dependiente de un parámetro, β , la inversa se convierte en una tarea cercana a lo imposible en algunos casos. En este punto es cuando se convierte casi en imprescindible el uso de la fórmula

$$\varphi_k(M^{-1}) = \varphi_m(M^{-1}) \varphi_{m-k}(M),$$

(véase (5.3)) con la que se puede obtener $\Phi_{Ch_k} [M(\xi_{n+1})]$ en función de $\varphi_i [M(\xi_{n+1})]$, sin necesidad de invertir la matriz (recuérdese que $\varphi_m(M^{-1})$ es simplemente $[\det M]^{-1}$).

- Cálculo del β que hace mínima la función anterior, utilizando la función *FindMinimum* del *Mathematica*. En esta se pueden modificar algunos de sus parámetros, como la precisión o el número de iteraciones que emplea en su algoritmo. De todas formas, es conveniente controlar aquellos casos en que no consigue su propósito, y dar a β en éstos un valor por defecto. Esto se puede hacer teniendo en cuenta que cuando la función no ha alcanzado su objetivo toma como valor la propia llamada a la función. Por ejemplo:

$$beta = FindMinimum[funcion, \{b, \{.2, .1\}, 0, 1\}]$$

verificará $beta[[2]] = \{b, \{.2, .1\}, 0, 1\}$ cuando no se haya conseguido alcanzar el mínimo. Se puede examinar pues esa condición y hacer en ese caso $\beta = 0.001$, por ejemplo.

- **Mínimo:** recibe como parámetros la función, **f**, la variable respecto de la que se quiere minimizar y el intervalo de búsqueda; y calcula el conjunto de mínimos de una función en el intervalo, así como el valor en esos puntos. Aunque el *Mathematica* dispone de la función *FindMinimum*, que busca los mínimos locales, ésta no se comporta bien

en los casos en que éstos no existan, devolviendo distintos mensajes de error. Se hace necesario, pues, readaptar esta función, evitando los mensajes no deseados, comprobando el valor de f en los extremos del intervalo y repitiendo la búsqueda a fin de encontrar puntos que produzcan el mismo valor para f , ya que *FindMinimum* en el caso de encontrar varios mínimos sólo devuelve el menor. Para el algoritmo utilizado este hecho no representa ningún problema, ya que en sucesivos pasos incorporaría los puntos no añadidos previamente. Aunque lo cierto es que esto retrasaría el resultado final, y en el caso de tratarse de un diseño simétrico no tomaría los mismos pesos para los puntos correspondientes (a no ser que se calculara el peso óptimo), teniendo que retocar el diseño en sucesivas etapas. Procedimiento:

- Establecer los parámetros de precisión de los resultados de la búsqueda y el número de subdivisiones que se harán del intervalo. Se establece como mínimo provisional el valor de la función en el extremo inferior.
 - Aplicar la función *FindMinimum* a cada uno de los subintervalos. Lo que se hace es retocarlos un poco para que se solapen ligeramente, a fin de que los extremos sean considerados también como puntos candidatos en alguna ocasión. Cuando se encuentra algún punto en el que la función toma un valor menor que la cota, se toma este valor como nueva cota y el punto como nuevo mínimo. Si el valor es igual que la cota, se añade el punto al conjunto de mínimos.
 - Comprobar el valor de la función en el extremo superior del intervalo inicial, punto que nunca ha sido candidato a mínimo en el proceso anterior, y proceder como en el punto previo.
- **Refinar** simplifica el diseño que se va obteniendo. La función se ejecuta tras un número adecuado de iteraciones (que por supuesto se puede definir según la naturaleza del problema) y recibe un parámetro de *cercanía* y otro de *peso mínimo*. El primero es el que le indica qué puntos están tan cercanos que los podemos considerar iguales. La manera en que se tratan estos casos es hacer la media de los puntos en cuestión, ponderada por sus pesos. Esto nos proporciona un nuevo punto entre los dos cuyo peso es la suma de los dos pesos individuales. El parámetro de *peso mínimo* indica qué puntos tienen peso despreciable, y por tanto

se pueden eliminar del diseño. El tratamiento que se ha dado a éstos es asignar su peso a un punto del diseño cercano a él, si existe alguno en estas condiciones, o repartirlo entre todos en caso contrario. Por supuesto hay que tener cuidado con el orden en que se hacen todas estas operaciones, así como de no eliminar demasiados puntos para evitar los diseños singulares. El procedimiento empleado ha sido el siguiente:

- Eliminar del diseño los puntos con pesos despreciables (según los parámetros iniciales), teniendo cuidado de conservar siempre al menos m puntos distintos en el soporte, para evitar que resulte un diseño singular. Tanto los eliminados como sus pesos se conservarán en **quitar**.
- Comparar los puntos de **quitar** con los que quedan en el diseño. Cuando uno de los eliminados, x , esté *cerca* (según el parámetro inicial) de uno de los del diseño actual se sustituye el punto del diseño por la media de los dos, ponderada por sus pesos. El peso del punto resultante será por supuesto la suma de los pesos de ambos, y el punto x se elimina de **quitar**.
- Actualmente quedan en **quitar** los puntos que no están cerca de ninguno de los del diseño. Pero podría ocurrir que en ocasiones hubiera varios cercanos entre sí y cuyos pesos acumulados superaran el peso despreciable. Para tratar adecuadamente esta situación comparamos los puntos de **quitar** entre sí, comprobando si están *cerca* como en el apartado anterior, y creando un nuevo conjunto **quitar**.
- Recorrer **quitar**, añadiendo a nuestro diseño los puntos que superen el peso despreciable (y, por supuesto, eliminándolos de **quitar**).
- Los puntos que quedan en **quitar** ya serán eliminados definitivamente. El peso conjunto de todos ellos, \mathbf{S} , va a ser repartido entre los puntos del diseño, proporcionalmente a sus pesos. Esto consiste simplemente en multiplicar el diseño por $1/(1-\mathbf{S})$.
- Finalmente tenemos un diseño cuyos puntos tienen todos peso no despreciable. Aún podría ocurrir que algunos de estos puntos fuesen *cercanos* entre sí. Por este motivo se recorre el diseño con dos contadores comparando todos los puntos, y creando un punto

nuevo como la media de los dos ponderada por sus pesos cuando ocurra ese caso.

- **Función principal**, que utiliza todas las funciones anteriores y algunas propias, como la que calcula el diseño en cada paso a partir del obtenido en el paso anterior y el punto que se tiene que incorporar, y cuyo procedimiento se describe a continuación:

- Hallar x_n , punto en el que se alcanza el mínimo de la varianza generalizada del diseño ξ_n ,

$$d(x, \xi_n) = f(x)^t \nabla \Phi_{Ch_k} [M(\xi_n)],$$

que a su vez hay que definir previamente como función. Para ello se utiliza la función **Mínimo**

- Calcular el β que minimiza $\Phi_{Ch_k} [M(\xi_{n+1})]$, donde

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

con la función descrita anteriormente

- Cambiar el diseño actual por ξ_{n+1} . Hay que tener en cuenta el caso en que x_n ya pertenezca al soporte del diseño, para el que se facilitan los cálculos.

Entonces el esquema de la función principal es el siguiente:

- Recibir los parámetros iniciales: el modelo (lineal), la variable en que viene dado, k , el intervalo soporte, el diseño inicial y el número de iteraciones (o alternativamente una cota para la condición de parada). Asimismo existen otra serie de parámetros auxiliares que se describirán más adelante.
- Definir las funciones auxiliares, como por ejemplo la varianza generalizada y la descrita anteriormente
- Crear un bucle donde vaya ejecutando esta función y modificando el diseño inicial. Cada cierto número de iteraciones (este número es uno de los parámetros configurables inicialmente), se llama a la función **Refinar** para depurar el diseño. Opcionalmente se va mostrando información escrita y gráfica del proceso. El número de iteraciones del bucle es uno de los parámetros iniciales. En el

caso de emplear condiciones de parada el bucle finaliza cuando se alcance la precisión exigida.

– Representación escrita y gráfica del diseño final.

- **Funciones informativas**, que van mostrando cada cierto número de iteraciones (definible) el diseño que se va obteniendo, el punto que se tiene que incorporar en cada caso, etc. A la hora de reflejar los resultados numéricamente resulta muy adecuada la ordenación de los puntos dentro del diseño. A este fin va encaminada la función **Ordenar**, que por supuesto es posible implementar por cualquiera de los múltiples procedimientos de ordenación conocidos. Al no tratarse de un gran número de puntos la elección del método no resulta de excesiva importancia. En el caso en que se quiera comprobar la bondad de los diseños que se van obteniendo respecto de algún criterio cuyo óptimo es conocido es posible utilizar la función **Eficiencia**.
- **Funciones gráficas**, que representan principalmente el diseño final con el grosor de los puntos proporcional a su peso. Opcionalmente se pueden mostrar los diseños que se van obteniendo en cada etapa, así como la varianza generalizada en cada paso.

Es importante en todo el proceso fijar un parámetro de *precisión* para evitar que puntos iguales sean tomados como diferentes. Esto puede ocurrir debido a que los resultados de los cálculos numéricos que emplea el programa utilizan una enorme cantidad de decimales, y por ello una mínima variación en un decimal despreciable provocaría que tomara como diferentes los puntos.

Bibliografía

- [1] Abdelbasit K. M. y Plackett R. L. (1983). Experimental design for binary data. *Journal of the american statistical association*, **78**, 381, 90-98.
- [2] Abramovitz M. y Stegun I. A. (1972). Handbook of mathematical functions. Dover Pub. Inc., New York, 773-801.
- [3] Albert A. (1972). Regression and the Moore-Penrose pseudoinverse. Academic Press. Ser. Mathematics in science and engineering, 94.
- [4] Ardanuy R., López Fidalgo J., Laycock P. J. y Wong W. K. (1999). When is an equally-weighted design D -optimal? *Ann. Inst. Statist. Math.*, **51**, 3, 531-540.
- [5] Ardanuy R. y López Fidalgo J. (1992). Optimal Design with Constraint Support. *Rev. Mat. Estat.*, Sao Paulo, 10, 193-205.
- [6] Ash A. y Hedayat A. (1978). An introduction to design optimality with an overview of the literature. *Commun. statist. -theor. meth.*, Ser. A, **7**, 14, 1295-1325.
- [7] Atkinson A. C. (1982). Developments in the design of experiments. *International Statistical review*, 50, 161-177.
- [8] Atkinson A.C., Chaloner K., Herzberg A.M. y Juritz J. (1992). Optimum experimental designs for properties of a compartmental model. *Biometrics*, **48**.
- [9] Atkinson A.C. y Cook R.D. (1995). D-optimum Designs for Heteroscedastic Linear Models. *Journal of the American Statistical Association*, 90, 429.

- [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*, **62**, 1, 57-70.
- [12] Atkinson A. C. y Fedorov V. V. (1975). Optimal design: Experiments for discriminating between several models. *Biometrika*, **62**, 2, 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, 465-473.
- [14] Atkinson A. C. y Hunter W. G. (1968). The design of experiments for parameter estimation. *Technometrics*, **10**, 2, 271-289.
- [15] Atwood C. L. (1969). Optimal and efficient designs of experiments. *The Annals of Mathematical Statistics*, **40**, 5, 1570-1602.
- [16] Atwood C. L. (1973). Sequences Converging to D-Optimal Designs of Experiments. *The Annals of Statistics*, **1**, 2, 342-352.
- [17] Atwood C. L. (1976). Convergent design sequences, for sufficiently regular optimality criteria. *The annals of statistics*, **4**, 6, 1124-1138.
- [18] Atwood C. L. (1980). Convergent design sequences, for sufficiently regular optimality criteria, II: singular case. *The annals of statistics*, **8**, 4, 894-912.
- [19] Bandemer H. (1980). Problems in foundation and use of optimal experimental design in regression models. *Math. Operationsforsch. Statistics.*, Ser. Statistics, **11**, 1, 89-113.
- [20] Bandemer H., Nather W. y Pilz J. (1987). Once more: Optimal experimental design for regression models (with discussion). *Statistics*, **18**, 2, 171-217.
- [21] Bates D. (1983). The derivative of $|X'X|$ and its uses. *Technometrics*, **25**, 4, 373-376.

- [22] 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*, **12**, 2, 77-90.
- [23] Bohning D. (1984). Use of reparameterization in nonlinear optimization with applications to statistics and optimal design. *Computational statistics quarterly*. **1**, 1, 29-43.
- [24] Box G. E. P. (1979). Some problems of statistics and everyday life. *Journal of the american statistical association*, **74**, 1-4.
- [25] 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*, **54**, 622-654.
- [26] Box G. E. P. y Hunter W. G. (1965). Sequential design of experiments for nonlinear models. *IBM Scientific computing symposium in statist.*, 113-137.
- [27] Box G. E. P. y Lucas H. L. (1959). Desing of experiments in non-linear situations. *Biometrika*, **46**, 77-90.
- [28] Box M. J. (1968a). The use of designed experiments in nonlinear model building. *The future of statist*, 241-268.
- [29] 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, **30**, 290-302.
- [30] 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, England.
- [31] Box M. J. (1970). Some experinces with a nonlinear experimental design criterion. *Technometrics*, **12**, 569-589.
- [32] Box M. J. y Draper N. R. (1971). Factorial designs, the $|X'X|$ criterion, and some related matters. *Technometrics*, **13**, 4, 731-742.
- [33] Brooks R. J. (1972). A decision theory approach to optimal regression designs. *Biometrika*, **59**, 3, 563.

- [34] 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.
- [35] Calvete H. I. (1983). Diseño óptimo de experimentos: Diseños marginalmente restringidos. Tesis doctoral, Facultad de Ciencias, Univ. de Zaragoza.
- [36] Chaloner K. y Verdinelli I. (1995). Bayesian Experimental Design: A Review. *Statist. Sci.* **10**, 3, 273-304.
- [37] Chernoff H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 24, 586-602.
- [38] Clarke G. M. (1980). *Statistics and experimental design*. Ed. Edward Arnold Ltd.
- [39] Cook R.D. y Wong W.K. (1994). On the Equivalence of Constrained and Compound Optimal Designs. *Journal of the American Statistical Association*, 89, 426.
- [40] Covey-Crump P. A. K. y Silvey S. D. (1970). Optimal regression designs with previous observations. *Biometrika*, **57**, 3, 551-566.
- [41] De La Garza A. (1954). Spacing of information in polynomial regression. *The Annals of Mathematical Statistics*, **25**, 123-130.
- [42] Dette H. (1993). On the Mixture of the D- and D1-optimality criterion in polynomial regression. *Journal of Statistical Planning and Inference*, **35**, 233-249.
- [43] Dette H. (1995). Optimal Design for Identifying the Degree of a Polynomial Regression. *Ann. Statist.* 4, 1248-1266.
- [44] Dette H. (1997). Designing Experiments with respect to 'Standardized' Optimality Criteria, *J.R. Statist. Soc.*, B **59**, 1, 97-110.
- [45] Dette H. (1997). A note on the uniform distribution on the arcsin points. *Metrika*, **46**, 1, 71-82.

- [46] Dette H., Haines L. M., Imhof L. (1999). Optimal designs for rational models and weighted polynomial regression. *Ann. Statist.* **27**, 4, 1272–1293.
- [47] Dette H., Heiligers B. y Studden W. J. (1995). Minimax Designs in Linear Regression Models. *Ann. Statist.* **1**, 30-40.
- [48] Dette H. y Sahn M. (1998). Minimax Optimal Designs in Nonlinear Regression Models. *Statistica Sinica* **8**, 4, 4249-4264.
- [49] Dette H. y Studden W. J. (1993). Minimax Designs in Linear Regression Models. *The Annals of Statistics*, **23**, 1, 30-40.
- [50] Dette H. y Wong W.K. (1999). E-optimal Design for the Michaelis-Menten Model. *Statist. Probab. Lett.* **44**, 4, 405–408.
- [51] Dodge, Fedorov y Wynn (Eds.) (1988). *Optimal design and analysis of experiments. First international conference workshop on optimal design and analysis experiments*. Elsevier science publishers B. V.
- [52] 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*, **54**, 662-665.
- [53] Ehrenfeld S. (1955). On the efficiency of experimental designs. *The annals of mathematical statistics*, **26**, 247-255.
- [54] Elfving G. (1952). Optimum allocation in linear regression theory. *The annals of mathematical statistics*, **23**, 255-262.
- [55] Elfving G. (1959). Designs of Linear Experiments. *Prob. and Stats. The Harold Cramér Volume*. Ed. Ulf Grenander Wiley, 58-74.
- [56] Federer W. T. (1980). Some recent results in experiment design with a bibliography - I. *International statistical review*, **48**, 357-368.
- [57] Fedorov V.V. (1972). *Theory of optimal experiments*. Department of statistics Purdue University.
- [58] Fedorov V. V. (1980). Convex design theory. *Math. Operationsforsh. Statist.*, Ser. statistics., **11**, 3, 403-413.

- [59] Fedorov V. V. (1988). Properties of optimal designs of regression experiments in singular cases. *Journal of soviet mathematics*, **40**, 2, pp 245-246.
- [60] Fedorov V. V. y Hackl P. (1997). *Model-Oriented Design of Experiments*. Lect. Notes in Statistics. Springer-Verlag New York, Inc.
- [61] Fedorov V. V. y Khabarov V. (1986). Duality of optimal designs for model discrimination and parameter estimation. *Biometrika*, **73**, 1, 183-190.
- [62] Ford I. (1976). Ph.D. Thesis. University of Glasgow.
- [63] Ford I., Titterington D. M. And Kitsos Ch. P. (1989). Recent Advances in Nonlinear Experimental Design. *Technometrics*, **31**, 1, 49-60.
- [64] 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*, **54**, 2, 569-583.
- [65] Frischmuth K. (1989). On locally D-optimal experimental design. *Rostock.math. kolloq.*, **36**, 51-64.
- [66] Galil Z. y Kiefer J. (1980). Time-and space- saving computer methods, related to Mitchell's DETMAX, for finding *D*-optimum designs. *Technometrics*, **22**, 3, 301-313.
- [67] Galil Z. y Kiefer J. (1983). Comparison of designs equivalent under one or two criteria. *Journal of statistical planning and inference*, 8, 103-116.
- [68] Gill E. P., Murray W. y Wright M. H. (1981). *Practical Optimization*. Academic Press.
- [69] Gladitz J. y Pliz J. (1982). Construction of optimal designs in random coefficient regression models. *Math. Operationsforsch. Statistics*, **13**, 3, 371-365.
- [70] Goldstein E. G. (1972). *Theory of convex programming*. American Mathematical Society providence. Rhode Island.
- [71] Graffke N. (1985). Directional derivatives of optimality criteria at singular matrices in convex design theory. *Statistics*, **16**, 3, 373-388.

- [72] Graybill F. A. (1983). *Matrices with applications in statistics*. Wadsworth.
- [73] Guest P. G. (1958). The spacing of observations in polynomial regression. *The annals of mathematical statistics*, **29**, 294-299.
- [74] Hardin R. H. y Sloane N. J. A. (1993). A new approach to the construction of optimal designs. *J. Statist. Plann. Inference*, **37**, 3, 339-369.
- [75] Hardin R. H. y Sloane N. J. A. (1994). Operating manual for Gosset: A general-purpose program for constructing experimental designs. *Mathematical Sciences Research Center. AT&T Bell Laboratories*, Murray Hill, New Jersey.
- [76] Hardy G., Littlewood J. E. y Polya G. (1989). *Inequalities*. Cambridge university press. New York, Port Chester, Melbourne y Sydney.
- [77] Healy M. J. R. (1986). *Matrices for statistics*. Oxford science publications.
- [78] Hedayat A. (1981). Study of optimality criteria in design of experiments. *Statistics and related topics*. North Holland Publishing Company, 39-56.
- [79] Herzberg A. M. (1982). The robust design of experiments: A review. *Serdica Bulgaricae mathematicae publicationes*, 8, 223-228.
- [80] Hill D. H. (1980). D-optimal designs for partially nonlinear regression models. *Technometrics*, **22**, 2, 275-276.
- [81] Hoel P. G. (1958). Efficiency problems in polynomial estimation. *The Annals of Mathematical Statistics*, 29, 1134-1145.
- [82] Hoel P. G. y Levine A. (1964). Optimal spacing and weighting in polynomial prediction. *The Annals of Mathematical Statistics*, 35, 1553-1560.
- [83] Holmes R. B. (1975). *Geometric functional analysis and its applications*. Springer, New York.
- [84] Imhof M., López Fidalgo J. y Wong W.K. (2000). Efficiencies of optimal approximate designs for small samples. *Statistica Nederlandica* (en prensa).

- [85] Jacroux M. (1983). On the MV-Optimality of Chemical Balance Weighing Designs. *Calcutta Statist. Assoc. Bull.* **32**, 143-151.
- [86] 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*, **15**, 205-225.
- [87] 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*, **20**, 201-214.
- [88] John J. A. y Quenovile M. H. (1977). *Experiments: design and analysis*. Ed. Charles Griffin and Company Ltd.
- [89] John R. C. St. y Draper N. R. (1975). D-optimality for regression designs: A review. *Technometrics*, **17**, 1, 15-23.
- [90] Karlin S. y Studden W. (1966). Optimal experimental designs. *The annals of mathematical statistics*, **37**, 783-810.
- [91] Kiefer J. (1959). Optimum experimental designs. *Journal royal statistical society*, Ser. B, **21**, 272-319.
- [92] Kiefer J. (1961). Optimum designs in regression problems, II. *The annals of mathematical statistics*, **32**, 298-325.
- [93] Kiefer J. (1974). General equivalence theory for optimum designs (Approximate theory). *The annals of statistics*, **2**, 5, 848-879.
- [94] Kiefer J. (1975a). Optimal design: Variation in structure and performance under change of criterion. *Biometrika*, **62**, 2, 277-288.
- [95] Kiefer J. (1975b). Construction and Optimality of Generalized Youden Designs. *J. N. Srivastava, ed., A Survey of Statistical Design and Linear Models*. North-Holand Publishing Company, 333-353.
- [96] Kiefer J. y Studden W. J. (1976). Optimal design for large degree polynomial regression. *The annals of statistics*, **4**, 6, 1113-1123.
- [97] Kiefer J. y Wolfowitz J. (1959). Optimum design in regression problems. *The Annals of Mathematical Statistics*, **30**, 271-294.

- [98] Kiefer J. y Wolfowitz J. (1960). The equivalence of two extremum problems. *Canad. J. Math.*, 12, 363-366.
- [99] Krafft O. (1983). A matrix optimization problem. *Elsevier Science publishing Co., Inc.*, 137-142.
- [100] Laycock P. J. (1972). Convex Loss Applied to Design in Regression Problems. *J. R. Statist. Soc.*, Series B, **34**, 148-170.
- [101] Laycock P. J. (1979). Multivariate optimal design. *Operations Research Verfahren*, Hain, Meisenheim **30**, 87-97.
- [102] Laycock P. J. y Rowley P. J. (1995). A method for generating and labelling all regular fractions or blocks for q^{n-m} -designs. *J. Roy. Statist. Soc. Ser. B* 57 (1995), **1**, 191-204.
- [103] Laycock P. J. y Silvey S. D. (1968). Optimal designs in regression problems with a general convex loss function. *Biometrika*, **55**, 1, 53-66.
- [104] López Fidalgo J. (1991). Algunos aspectos notables del Diseño Óptimo de Experimentos. Tesis Doctoral. Salamanca.
- [105] López Fidalgo J. (1992). Minimizing the Largest of the Parameters Variances. $V(\beta)$ -optimality. Müller W.G., Wynn H.P. and Zhigljavsky A.A. (eds). *Model-Oriented Data Analysis*, Physica-Verlag, Heidelberg, 71-79.
- [106] López Fidalgo J., Ardanuy Albajar R. y Rodríguez Díaz J.M. (1997). Convex Seeking of Optimal Designs. *Proceedings of the Second World Conference of The International Association for Statistical Computing*, Wegman y Azen Eds, 125-29. Los Angeles.
- [107] López Fidalgo J. y Rodríguez Díaz J.M. (1997). Algunos Resultados sobre el Orden de Loewner de las Matrices de Información, *Actas del XXIII Congreso Nacional de Estadística e Investigación Operativa*, 30.5-30.6. Valencia.
- [108] López Fidalgo J. y Rodríguez Díaz J.M. (1998). Characteristic Polynomial Criteria in Optimal Experimental Design, *MODA 5 Advanced in Model-Oriented Data Analysis and Experimental Design*, Physica-Verlag, 31-38.

- [109] López Fidalgo J., Torsney B. y Ardanuy R. (1998). MV-Optimization in Weighted Linear Regression, *MODA 5- Advances in Model-Oriented Data Analysis and Experimental Design*, Physica-Verlag, Heidelberg, 39-50.
- [110] López Fidalgo J. y Wong W.K. (2000). A comparative study of MV- and SMV-optimal designs for binary response models, *Advances in Stochastic Simulation Methods*, N. Balakrishnan, V.B. Melas, S. Ermakov, editors Birkhauser, Boston. Basel, Berlin, 135-152.
- [111] Minkin S. (1987). Optimal designs for binary data. *Journal of the american statistical association*, **82**, 4000, 1098-1103.
- [112] Mitchell T. J. (1974). An algorithm for the construction of "D-optimal" experimental designs. *Technometrics*, **16**, 2, 203-210.
- [113] Nalimov V. V., Golikova T. I. y Mikeshina N. G. (1970). On practical use of the concept of D-optimality. *Technometrics*, **12**, 4.
- [114] Nigam A. K., Gupta S. C. y Gupta S. (1983). A new algorithm for extreme vertices designs for linear mixture models. *Technometrics*, **25**, 4, 367-371.
- [115] Ortiz Rodríguez I. M. (1997). Algunas Aportaciones al Diseño Optimo de Experimentos para Modelos Lineales de Regresión. Tesis Doctoral. Departamento de Estadística y Matemática Aplicada de la Universidad de Almería.
- [116] Ortiz I., Martínez I. y Rodríguez C. (1999). *Diseño Optimo para Caracterizar la Relación Producción-fertilizante*. VII Conferencia Española de Biometría.
- [117] Papakiriakis P. A. (1978). Optimal experimental design in econometrics. The time series problem. *Journal of econometrics*, 7, 351-372.
- [118] Pázman A. (1974). A Convergence Theorem in the Theory of D-Optimum Experimental Designs. *The Annals of Statistics*, **2**, 1, 216-218.
- [119] Pázman A. (1980a). Singular experimental designs (Standard and Hilbert-space approaches). *Math. operationsforsch. statist.*, Ser. statistics, **11**, 1, 137-149.

- [120] Pázman A. (1980b). Some features of the optimal designs theory -A survey. *Math. Operationsforsch. Statist.*, Ser. statistics, **11**, 3, 415-446.
- [121] Pázman A. (1986). *Foundations of optimum experimental design*. D. Reidel publishing company. Dordrecht.
- [122] Peña Sánchez de Rivera D. (1995). *Estadística: Modelos y métodos*, Vol 1 y 2. Alianza Universidad Textos.
- [123] Pukelsheim F. (1980). On linear regression designs wich maximize information. *Journal of statistical planning and inference*, **4**, 339-346.
- [124] Pukelsheim F. (1981). On c-Optimal Design measures. *Math. Operationsforsch. Statist.*, Ser. Statistics, **12**, 1, 13-20.
- [125] Pukelsheim F. y Torsney B. (1991). Optimal Weights for Experimental Designs on Linearly Independent Support Points. *The Annals of Statistics*, **19**, 3, 1614-1625.
- [126] Pukelsheim F. (1993). *Optimal Design of Experiments*. John Wiley and Sons. New York.
- [127] Rao C. R. (1962). A note on a generalized inverse of a matrix with applications to problems in matematical statistics. *Journal royal statistical society*, Ser. B, **24**, 152-158.
- [128] Rasch D. y Darius P. (1995). Computer Aided Design of Experiments. *Computational Statistics* (1995). Physica-Verlag, Heidelberg.
- [129] Rasch D. y Hendrix E. (1997). Replicationfree Optimal Designs in Regression Analysis. *Computational Statistics (1997)*, 12, 19-52. Physica-Verlag, Heidelberg.
- [130] Rodríguez Díaz J.M. y López Fidalgo J. (1998). Un Estudio del Polinomio Característico de la Matriz de Información, *Actas del XXIV Congreso Nacional de Estadística e Investigación Operativa*, Almería, 101-2.
- [131] Rodríguez Díaz J.M. y López Fidalgo J. (2000). Optimal Characteristic Designs for Polynomial Models, *Optimum Design 2000. Atkinson, Zhigljavsky y Bogacka (Eds.). Kluwert. Cardiff, 2000*, 123-130.

- [132] Rodríguez Torreblanca C. y Ortiz Rodríguez I.M. (1999) *Diseño óptimo de experimentos para modelos de regresión*. Universidad de Almería, Servicio de publicaciones. Almería.
- [133] Rowley P. J. y Laycock P. J. (1991). Minimum Aberration Designs. *Technical Report 201*, Department of Mathematics, UMIST, Manchester.
- [134] Sanchez León J.G. (1998). Variables estocásticas en modelos compartimentales. Tesis Doctoral. Salamanca.
- [135] Schwabe R. (1996). Optimum Designs for Multi-Factor Models. *Lect. Notes in Statistics*. Springer-Verlag New York, Inc.
- [136] Shah K.R. y Sinha B.K. (1989). *Theory of Optimal Designs*. Lect. Notes in Statistics. Springer-Verlag Berlin Heidelberg.
- [137] Shisha O. (1967). *Inequalities*. Academic Press.
- [138] Silvey S. D. (1974). Some aspects of the theory of optimal linear regression design with a general concave criterion function. *Techn. rep.*, 75, Ser. 2. Department of statistics, Princenton university.
- [139] Silvey S.D. (1980). *Optimal design*. Chapman and Hall.
- [140] Silvey S. D. y Titterington D. M. (1973). A geometric approach to optimal design theory. *Biometrika*, **60**, 1, 21-32.
- [141] Smith K. (1918). On the standard deviations of adjusted and interpolates values of an observed polynomial functions and its constansts and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12, 1-85.
- [142] Snee R. D. y Marquardt W. (1974). Extreme vertices designs for linear mixture models. *Technometrics*, **16**, 3, 399-408.
- [143] Spruill M.C. (1990). Good Designs for Testing the Degree of a Polynomial Mean. *Sankhya*, **52**, Ser. B, 1, 67-74.
- [144] Steinberg D. M. y Hunter W. G. (1984). Experimental design: Review and coment. *Technometrics*, **26**, 2, 71-130.

- [145] Studden W.J (1980). D_s -optimal designs for polynomial regression using continued fractions. *The Annals of Statistics*, **8**, 5, 1132-1141.
- [146] Szego G. (1939). *Orthogonal Polynomials*. American Mathematical Society providence. Rhode Island.
- [147] Titterington D.M. (1975). Optimal design: Some geometrical aspects of D -optimality. *Biometrika*, **62**, 2, 313-320.
- [148] 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*.
- [149] Torsney B. (1981). A moment inequality and monotonicity of an algorithm. *Lecture Notes in Economics and Mathematical Systems. Semi-infinite Programming and Applications (Proceedings)*.
- [150] Torsney B. (1986). Moment inequalities via optimal design theory. *Linear Algebra and its Applications*, **82**, 237-253.
- [151] 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, 121-129.
- [152] Torsney B. y López Fidalgo J. (1995). MV-Optimization in Simple Linear regression. *Kitsos C.P. and Müller W.G. (eds). MODA 4: Advances in Model-Oriented Data Analysis*, Physica-Verlag, Heidelberg, 57-69
- [153] Wald A. (1943). On the efficient design of statistical investigations. *The Annals of Mathematical Statistics*, **14**, 134-140.
- [154] Welch W. J. (1982). Branch-and-bound search for experimental design based on D -optimality and other criteria. *Technometrics*, **24**, 1, 41-48.
- [155] White L. V. (1973). An extension of the general equivalence theorem to non linear models. *Biometrika*, **60**, 2, 345-348.
- [156] Whittle P. (1973). Some general points in the theory of optimal experimental design. *Journal royal statistical society, Ser. B*, 1, 123-130.

- [157] Wong W. K. (1995). On the Equivalence of D- and G-optimal Designs in Heteroscedastic Models. *Statistics and Probability Letters*, 25, 317-321.
- [158] Wu Ch -F. (1978). Some algorithmic aspects of the theory of optimal designs. *The annals of statistics*, 6, 6, 1286-1301.
- [159] Wu Ch.-F. y Wynn H. P. (1978). The convergence of general step-length algorithms for regular optimum design criteria. *The annals of statistics*, 6, 6, 1273-1285.
- [160] Wynn H. P. (1970). The sequential generation of D-optimum experimental designs. *The annals of mathematical statistic*, 41, 5, 1655-1664.
- [161] Wynn H.P. (1972). Results in the theory and construction of D-optimum experimental design. *Journal royal statistical society*, Ser. B, B, 34, 133-147.
- [162] 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, 571-579.

Índice

- ς_k , **100**
- A^+ , 33
- A^- , 33
- $C_g()$, **84**
- Cov*, covarianza, 7
- $Cov_\xi(g_1, g_2)$, 34
- $C_p()$, **84**
- D_{ij} , **88**
- D_l , **87**
- \mathcal{E} , **57**
- \mathcal{E}_ξ , **58**
- Φ_{Ch_k} , **97**
- $\Phi(\xi)(= \Phi[M(\xi)])$, 42
- $\Phi_{p,k}$, **103**
- Ξ , 30
- Ξ_Φ^* , **40**
- $\mathcal{L}(\mathcal{M})$, **40**
- \mathcal{M} , 31
- \mathcal{M}_+ , **40**
- \mathcal{M}_Φ , **40**
- $\mathcal{M}()$, 29
- $[M]_{i_1, \dots, i_k}$, **92**
- $M(\xi)$ =mat. de inf. de ξ , 30
- $[M]_k$, **92**
- $\mathcal{N}()$, 29
- $NND(m)$, **38**
- $PD(m)$, **39**
- \mathbb{R} , recta real
- \mathcal{R} , **55**
- Σ_ε , 11
- S_r , 32
- Θ , espacio de estados, 26
- U_Φ , 59, 60
- φ_k , **92**
- $Var_\xi g$, 34
- X_ξ , soporte de ξ , **27**
- β_n , 69
- conv*=cierre convexo, 36
- $d(x, \xi)$, **68**
- eff*, **42**
- \sim , 83
- $E[]$, esperanza
- η , superficie respuesta, 26
- $f(x_1, \dots, x_n)$, f. de densidad, 8
- g , funcional lineal, **28**
- γ , 70
- ∇ , gradiente
- ξ , diseño, **27**
- $\lambda_1, \dots, \lambda_m$, 35
- m =núm de parám del mod, 26
- μ , media, 7
- ∇ , gradiente
- $\|A\|$, **43**
- $|c|_1$, **55**
- $|c|_2$, **55**
- $|c|_{\mathcal{R}}$, **55**
- $\langle A, B \rangle$, **43**
- ψ , **62**
- ρ , correlación, 7
- σ^2 , varianza, 7

- θ^* , 7
- $\hat{\theta}$, 8
- θ , estado, 25
- ε , 8, 11
- X , espacio de diseño, 25
- (X, Θ, σ) , 26
- adding*, procedimiento, **73**
- aleatorización, 16
- algoritmo
 - acelerado, **70**
 - ascendente, **69**
 - combinado, **74**
 - de paso armónico, **68**
 - local, 75
 - para la obtención del óptimo, 22
 - para un X finito, 71
- A -optimización, 20, **45**
- autovalor, 47, 49
- autovector, 47
- base
 - cambio de, 93
 - ortogonal, 93
- biocinética, 121
- bloque, *ver* variable bloque
- BLUE, **29**
- branch and bound*, proced., 73
- Caratheodory, teorema de, 35
- cierre convexo, **35**
- coeficientes característicos, **92**
- compartimentos, 122
- concavidad, **38**
- convexidad, **38**
- c -optimización, **50**
- correlación, **7**
- covarianza, **7**
- generalizada, **34**
- criterio
 - alfabético, 21
 - de optimización, **40**
 - parcial, 41
 - singular, *ver* c. o. parcial
 - estandarizado, **53**
 - minimax, **46**
- cuadrado
 - grecolatino, *ver* mod. de ANOVA
 - latino, *ver* modelos de ANOVA
- diagonalizar, 93
- diseño
 - óptimo, **19, 40**
 - aproximado, **27, 30**
 - asintótico, *ver* aproximado
 - comparable, **84, 85, 86, 88, 89**
 - continuo, **27**
 - de experimentos, 7, 19
 - de tamaño fijo, *ver* exacto
 - equivalente, **83**
 - exacto, **27, 29**
 - localmente óptimo, **18**
 - mejor, **83**
 - no peor, **83**
 - sin replicación, 72
 - unipuntual, 42, 85, 87
- distribución
 - a posteriori, **7**
 - a priori, **7**
 - de probabilidad, 6
 - final, **7**
 - inicial, **7**
- DN -optimización, 22
- dominio experimental, *ver* espacio de diseño
- D -optimización, 20, **44**

- dropping*, procedimiento, **73**
- efectos fijos, **15**
- eficiencia
 - de un diseño, **42**
- elipsoide de confianza, 22, 91
- E*-optimización, 21, **47**
- equivalencia, teorema, 21
- error
 - cuadrático medio, 8
- espacio
 - de diseño, 5, **25**
 - paramétrico, 7
- esperanza, **6**
- estado, 5, 11, **25**
- estimación, 7
- estimador
 - Bayes, **8**
 - centrado, **8**, 28
 - lineal, **28**
 - máximo verosímil, **8**, 13
 - mínimo cuadrático, 13
- experimento científico, 1, 5
- extensión, procedimiento, **74**
- factor, **15**
- farmacocinética, 121
- Φ_p -optimización, **49**
- función
 - critério, 40, **40**
 - clases de, 43
 - convexa, **59**
 - estrictamente creciente, **41**
 - homogénea, **42**
 - de pérdida, **7**
 - de regresión, 26
 - de sensibilidad, 62
 - de utilidad, 18
 - de verosimilitud, 8
- funcional lineal, **28**
- estimable, **29**
- funciones
 - características, **97**
 - simétricas, 92
 - simétricas elementales, **92**
- g*-inversa, 33
- GN*-optimización, 22
- G*-optimización, 21, **45**
- gradiente, **43**
- Hadamard, desigualdad de, 93
- heterocedasticidad, 11
- homogeneidad (criterios), **42**
- homogeneidad positiva, **39**
- intercambio, procedimiento, 74
- inversa generalizada, **33**
 - de Penrose, **33**
- isotonicidad, **38**
- Legendre, polinomios de, 112
- Loewner, orden de, **83**
- L*-optimización, **48**
- L_p*-optimización, **49**
- mínimos cuadrados, 9
- matriz
 - de covarianzas, 9
 - de dispersión, *ver* m. de covarianzas
 - de información, 13, 20
 - de Fisher, 17
 - de un diseño aproximado, **30**
 - de un diseño exacto, **29**
 - regular, *ver* no singular
 - singular, **12**

- media, **6**
- medida de probabilidad, **27**
- minimax, criterio, **46**
- mixturas de modelos, **23**
- modelo
 - compartimental, **122**
 - abierto, **123**
 - cerrado, **123**
 - con emparejamiento lineal, **14**
 - de ANOVA, **15**
 - con efectos aleatorios, **17**
 - cuadrado grecolatino, **17**
 - cuadrado latino, **16**
 - en bloques aleatorizados, **16**
 - factorial, **16**
 - de Michaelis-Menten, **23**
 - de observaciones directas, **13**
 - de observaciones indirectas, **13**
 - bajo restricciones lineales, **14**
 - de regresión, **8, 13**
 - lineal, **11**
 - spline, **14**
 - exponencial, **120**
 - lineal, **9, 26**
 - multifactorial, **23**
 - no lineal, **17, 22**
 - polinómico, **111**
- muestra, **1, 6**
- MV*-optimización, **48**

- N =tamaño muestral, **6**
- norma, **43**
 - de la traza, **43**

- parámetro, **6**
- paso armónico, algoritmo de, **68**
- población, **1, 6**
- polinomio
 - homogéneo, **99**
 - polinomio característico, **91, 93**
 - polinomio característico, **92**

- reducción, procedimiento, **74**
- regresión
 - logística, **23**

- semidefinido positiva, **31**
- sensibilidad, función de, **62**
- simetría, **22**
- soporte, **27**
- soporte arcoseno, **119**
- subconjunto, procedimiento, **74**
- superaditividad, **39**
- superficie respuesta, **26**

- tratamiento, *ver* var. tratamiento

- variable
 - bajo control, **16**
 - bloque, **16**
 - explicativa, **15**
 - fuera de control, **16**
 - respuesta, **15**
 - tratamiento, **16**
- varianza, **7**
 - generalizada, **34**
- verosimilitud, función de, **8**
- V*-optimización, **50**