



Diseños óptimos en la presencia de efectos de bloques aleatorios

Eddie Edinson Rodríguez Bossio

Tesis De Maestría

Barranquilla, Diciembre 2015



SOBRE UNA NUEVA CLASE DE POLINOMIOS TIPO APOSTOL GENERALIZADOS Y ALGUNAS PROPIEDADES

Eddie Edinson Rodríguez Bossio

Trabajo de Maestría

Barranquilla, Diciembre 2015



Ejemplo de Tesis

Trabajo de grado presentado para optar al grado Maestría en Matemática
para la Universidad del Atlántico (UA)

Este trabajo de investigación ha sido realizada bajo la dirección de
Dr. Director [†]

[†] Departamento de Matemática,
Universidad del Atlántico (UA)

Trabajo de investigación en Matemática,
Universidad del Atlántico (UA)

Dr. advisors:
DR. [†]

[†] Departamento de Matemática
Universidad del Atlántico (UA)

Barranquilla, Diciembre 2015



UNIVERSIDAD DEL ATLÁNTICO
PROGRAMA DE MATEMÁTICAS
MAESTRÍA EN MATEMÁTICAS

SOBRE UNA NUEVA CLASE DE POLINOMIOS TIPO APOSTOL GENERALIZADOS Y ALGUNAS PROPIEDADES

Estudiante: Pedro Hernández Llanos
Carnet: 1048206189

Este Trabajo Especial de Grado ha sido aprobado en nombre de la Universidad del Atlántico por el siguiente jurado examinador:

Dr. Boris Lora Castro
Universidad del Atlántico - Colombia

Dra. Yamileth Quintana Mato
Universidad Simón Bolívar - Venezuela

Alejandro Urieles Guerrero.
Tutor

07 de Marzo de 2014

AGRADECIMIENTOS

A mis padres y a Diana Patricia, Adriana y Diana Lucía .

El estudio de los polinomios tipo Apostol comenzó en 1951 y fue gracias a T. M. Apostol [?] quien propuso un nuevo conjunto de polinomios $\mathfrak{B}_n(x; \lambda)$, $\lambda \in \mathbb{C}$ que resultan ser una generalización de los números de Bernoulli. Estos nuevos polinomios están relacionados con la función Zeta de Lerch $\phi(x, \alpha, s)$ y son llamados polinomios Apostol-Bernoulli.

En 1985 el matemático francés Todorov [?] estudia una generalización de los polinomios $B_n^{(\alpha)}$ donde α es un parámetro real o complejo. Tres años después, Todorov en colaboración con Srivastava [?] prueba una fórmula explícita para los polinomios de Bernoulli generalizados $B_n^{(\alpha)}(x)$, $\alpha \in \mathbb{C}$:

$$B_n^{(\alpha)} = \sum_{k=0}^n \binom{n}{k} \frac{k!}{(2k)!} \sum_{j=0}^k (-1)^j \binom{k}{j} j^{2k} (x+j)^{n-k} \times F[k-n, k-\alpha; 2k+1; j/(x+j)] \quad (1)$$

De las ideas del Apostol y Todorov, muchos matemáticos comenzaron la tarea de extender estas ideas a otros polinomios con estructuras análogas a polinomios de Bernoulli como los polinomios de Euler y polinomios de Genocchi. A continuación se muestran las obras más destacadas en este campo:

En 2005 Luo y Srivastava [?] establecieron definiciones análogas a las de Apostol para los números y los también llamados polinomios Apostol-Bernoulli de orden superior $\mathfrak{B}_n^{(\alpha)}(x; \lambda)$. En el mismo año descubrieron una relación entre polinomios de Apostol-Euler generalizados y los polinomios Apostol-Bernoulli generalizados (see [?]).

Luo en 2006 [?] proporciona definiciones análogas a las de Apostol para los también llamados

números $E_n^{(\alpha)}(\lambda)$ y polinomios Apostol-Euler de orden superior $E_n^{(\alpha)}(x; \lambda)$.

En 2007 W.Wang, T. Wang y Jia [?] publican algunos resultados sobre los polinomios de Apostol-Bernoulli generalizados y los polinomios Apostol-Euler generalizados, específicamente establecen relaciones entre los dos polinomios y hacen un estudio en sumas de productos de estos polinomios.

En 2010 Kurt [?], introduce una generalización de los polinomios de Bernoulli $B_n^{[m-1, \alpha]}(x)$. En este trabajo se dan algunas relaciones de recurrencia de los nuevos polinomios y números de Bernoulli generalizados. Además, se muestra una relación entre polinomios los polinomios 2D-Bernoulli y los polinomios de Hermite Kamp 'e de Feriet.

En 2011 Tremblay, Gaboury, y Fugère [?], dan una generalización de la definición de Kurt $\mathfrak{B}_n^{[m-1, \alpha]}(x; \lambda)$, estudian algunas propiedades y establecen un analogo al teorema de la adición de Srivastava-Pintér. El mismo año los autores [?] estudian una nueva clase de polinomios Apostol-Bernoulli generalizados $\mathfrak{B}_n^{[m-1, \alpha]}(x, b, c; \lambda)$. En 2012, estos autores [?] introducen e investigan dos nuevas clases de polinomios Apostol-Euler generalizados $\mathfrak{E}_n^{[m-1, \alpha]}(x, b, c; \lambda)$ y polinomios Apostol-Genocchi generalizados $\mathfrak{G}_n^{[m-1, \alpha]}(x, b, c; \lambda)$, obtienen propiedades para las nuevas clases de polinomios Apostol-Euler generalizados y Apostol-Genocchi generalizados, además dan una extensión y un análogo del teorema de la adición de Srivastava-Pintér.

En 2013 Cheng, Cai y Luo [?] contruyen una extension para los polinomios de Apostol-Euler generalizados y los polinomios de Apostol-Bernoulli generalizados (notados por $E_{n, \lambda}^{[m-1, \alpha]}(x)$ y $B_{n, \lambda}^{[m-1, \alpha]}(x)$ respectivamente), específicamente estudian las propiedades de dicha extension sólo para los polinomios Apostol-Euler generalizados y derivan relaciones de estos con los números de Stirling de segunda especie, polinomios de Jacobi, polinomios de Laguerre, polinomios de Hermite and los polinomios de Bernoulli generalizados.

En el mismo año Luo, en colaboración con Lu [?] introduce los polinomios tipo Apostol generalizados $\mathcal{F}_n^{(\alpha)}(x; \lambda; u, v)$ con $(\alpha \in_0, \lambda, u, v \in \mathbb{C})$ y estudia sus propiedades. Estos polinomios generan todos los polinomios Apostol conocidos, así:

$$\mathcal{B}_n^{(\alpha)}(x; \lambda) = (-1)^\alpha \mathcal{F}_n^{(\alpha)}(x; -\lambda; 0, 1), \quad (2)$$

$$\mathcal{E}_n^{(\alpha)}(x; \lambda) = \mathcal{F}_n^{(\alpha)}(x; \lambda; 1, 0), \quad (3)$$

$$\mathcal{G}_n^{(\alpha)}(x; \lambda) = \mathcal{F}_n^{(\alpha)}(x; \lambda; 1, 1). \quad (4)$$

Entre los resultados más recientes podemos citar los de Lu, Luo y Xiang [?] siguiendo sus trabajos sobre los polinomios tipo Apostol generalizados, estudian algunos resultados para aquellos de estos polinomios que están asociados con el álgebra umbral. En este caso introducen una unificación de los polinomios tipo apostol notado por $\mathcal{Y}_{n,\beta}^{(\alpha)}(x; k; a, b)$, la cuál está relacionada con $\mathcal{F}_n^{(\alpha)}(x; \lambda; u, v)$ de la siguiente manera:

$$\mathcal{Y}_{n,\beta}^{(\alpha)}(x; k; a, b) = -\frac{1}{a^b} \mathcal{F}_n^{(1)}\left(x; -\left(\frac{\beta}{a}\right)^b; 1-k, k\right) \quad (5)$$

y

$$\mathcal{Y}_{n,\beta}^{(\alpha)}(x; k; a, b) = (-1)^\alpha \frac{1}{a^{b\alpha}} \mathcal{F}_n^{(\alpha)}\left(x; -\left(\frac{\beta}{a}\right)^b; 1-k, k\right). \quad (6)$$

Aunque los estudios correspondientes a los polinomios de tipo Apostol generalizados son los más recientes, en los últimos meses ha atraído una considerable atención, en especial los que están relacionados con el álgebra umbral y el estudio de sus propiedades.

¿Es posible generalizar $\mathcal{F}_n^{(\alpha)}(x; \lambda; u, v)$? ¿Qué relación guarda $\mathcal{F}_n^{(\alpha)}(x; \lambda; u, v)$ con otras familias de polinomios y números? Por lo que sabemos hasta la fecha, para resolver completamente estos problemas, es decir, dados los polinomios $\mathcal{F}_n^{(\alpha)}(x; \lambda; u, v)$ lograr obtener una expresión más general y establecer fórmulas de conexión con otras familias de polinomios y números se necesitan una gran variedad de herramientas entre ellas las ideas de [?] y [?].

En este trabajo se introducirá una nueva clase de polinomios tipo Apostol generalizados denotada por $\mathcal{Q}_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$, la cual es extensión natural de $\mathcal{F}_n^{(\alpha)}(x; \lambda; u, v)$. Así como también, se estudiarán sus propiedades básicas, la fórmula de recurrencia, la ecuación diferencial que satisfacen, se establecerán fórmulas de conexión de $\mathcal{Q}_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$ con los polinomios ortogonales clásicos, con los polinomios de Genocchi, Charlier, Bessel, Bernoulli generalizados y los números de Stirling de segunda clase. Finalmente se introducirá una nueva clase de polinomios tipo Apostol generalizados basados en los polinomios de Hermite de dos variables. La importancia de $\mathcal{Q}_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$ radica en que dichos polinomios generan a todos polinomios de Apostol conocidos, sus generalizaciones y las correspondientes extensiones de dichas generalizaciones. Teniendo en cuenta lo anterior y observando las últimas investigaciones vemos que muchos otros aspectos, como un estudio detallado de las propiedades de $\mathcal{Q}_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$ y su relación con otros polinomios y números, están pendientes de realización. Además, al ser $\mathcal{Q}_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$ una generalización de $\mathcal{F}_n^{(\alpha)}(x; \lambda; u, v)$ se

garantiza que el estudio de esta nueva clase de polinomios tipo Apostol generalizados es más general que el estudio hasta ahora conocido.

El trabajo en su totalidad, está constituido por cuatro capítulos que describiremos a continuación.

El capítulo I es de carácter introductorio. Contiene conceptos y resultados generales acerca de la teoría de sistemas ortogonales de funciones que se utilizarán a lo largo del trabajo. Haremos énfasis en los sistemas de polinomios ortogonales, en particular en los sistemas clásicos y los polinomios de Charlier y Bessel. También presentamos resultados relacionados con los polinomios de Bernoulli, Euler, Genocchi y las distintas generalizaciones que se desarrollaron hasta llegar a los polinomios de Apostol. Además se introducirán los números de Stirling de primera y segunda clase que más adelante necesitaremos para abordar el problema de investigación.

En el capítulo II, hacemos un resumen de las distintas generalizaciones de los polinomios de Apostol-Bernoulli, polinomios de Apostol-Euler, polinomios de Apostol-Genocchi hasta llegar a los polinomios tipo Apostol generalizados, al igual que explicamos algunos de los resultados más importantes de estos polinomios que serán de gran utilidad en el siguiente capítulo.

En el capítulo III, la sección 3.1 es dedicada al establecer las propiedades básicas de la nueva clase de polinomios tipo Apostol generalizados $Q_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$. En la sección 3.2 estudiamos la fórmula de recurrencia y ecuación diferencial para $Q_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$. La sección 3.3 por aplicación de algunos métodos de la sección anterior está dedicada a presentar fórmulas de conexión entre los polinomios $Q_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$ y los polinomios de Genocchi, los polinomios de Jacobi, los polinomios de Hermite, los polinomios de Laguerre, los polinomios de Charlier, los polinomios de Bessel, los polinomios de Bernoulli generalizados $B_n^{[m-1]}(x)$ y los números de Stirling de segunda especie. En la sección 3.4 introducimos una nueva clase de polinomios tipo Apostol generalizados basados en los polinomios de Hermite de dos variables al igual que indicamos algunas propiedades básicas.

Por último, en el capítulo IV exponemos las conclusiones y planteamos varios problemas que pueden conducir a futuras investigaciones. La tesis la completamos con una bibliografía relativa a los temas tratados en el trabajo y basada en artículos recientes y textos clásicos de indudable interés.

Contenido

APROBACIÓN DEL JURADO	v
Introducción	ix
Contenido	xiii
Lista de Tablas	xv
Lista de Figuras	xvii
Agradecimientos	xix
Abstract	xxi
Resumen	xxiii
I Preliminares	1
1 Preliminares	3
1.1 La ecuación de regresión	3
1.2 Estimación de mínimos cuadrados	5
1.3 Propiedades de los estimadores de mínimos cuadrados	16

1.4	Teorema de Gauss - Markov	17
1.5	Diseños experimentales y matrices de información	18
1.6	Criterios de optimalidad	21
1.6.1	D – criterio	22
1.6.2	G – criterio	22
1.6.3	D_s – criterio	23
1.6.4	Teorema de equivalencia	24
1.7	Diseño Factorial de Dos Factores Aleatorios	27
 II		29
2	Diseños óptimos en la presencia de efectos de bloques aleatorios	31
2.1	Aspectos del diseño	35
3	Conclusiones y trabajos futuros	39
REFERENCIAS BIBLIOGRÁFICAS		41
 III Appendices		43
Index		45

Lista de Tablas

1.1	Experimento turbina de gas descrito por Myers y Montgomery (1995).	11
1.2	Forma codificada del experimento de turbinas de gas.	12

Lista de Figuras

Agradecimientos

Doy las gracias a Dios que es el principal responsable de que finalizara de manera exitosa este trabajo. Es muy difícil expresar con palabras la inmensa gratitud que le debo al profesor Alejandro Urieles Guerrero, ya que durante estos casi tres años de mi permanencia en la Universidad del Atlántico, además de mostrarme el camino de la investigación en los polinomios ortogonales y tenerme una paciencia infinita, también me ayudó a encontrar el camino a la investigación de los polinomios tipo Apostol. El llegar a un feliz término en esta meta que me fijé hace mucho tiempo, es, en un gran porcentaje, gracias a él. Mil gracias Profesor Urieles.

También debo agradecer al departamento de Matemáticas de la Universidad del Atlántico, su apoyo no sólo económico, mediante la beca de la que he gozado durante los últimos años, sino la acogida que me brindaron cada uno de los integrantes de este agradable departamento. Particularmente, quiero resaltar el apoyo que me brindó el profesor Jorge Rodríguez Contreras, quien, como director de departamento, colaboró para que pudiera realizar muchas de las actividades académicas a las que tuve la suerte de asistir y quien además en los inicios de este proyecto, me proporcionó ayuda y consejo.

Quiero resaltar el sacrificio y dedicación de mis padres, lo que me ha permitido ser lo que soy. Finalmente, quiero decirles a Diana Patricia, Adriana y Diana Lucía que este proyecto no habría sido posible sin su apoyo y compañía, y que este logro también es de ustedes.

Pedro Hernández

The sequences are very versatile data structures. In a straightforward manner, a sequence of symbols can store any type of information. Systematic analysis of sequences is a very rich area of algorithmics, with lots of successful applications. The comparison by sequence alignment is a very powerful analysis tool. Dynamic programming is one of the most popular and efficient approaches to align two sequences. However, despite their utility, alignments are not always the best option for characterizing the function of two sequences. Sequences often encode information in different levels of organization (meta-information). In these cases, direct sequence comparison is not able to unveil those higher-order structures that can actually explain the relationship between the sequences.

We have contributed with the work presented here to improve the way in which two sequences can be compared, developing a new family of algorithms that align high level information encoded in biological sequences (meta-alignment). Initially, we have redesigned an existent algorithm, based in dynamic programming, to align two sequences of meta-information, introducing later several improvements for a better performance. Next, we have developed a multiple meta-alignment algorithm, by combining the general algorithm with the progressive schema. In addition, we have studied the properties of the resulting meta-alignments, modifying the algorithm to identify non-collinear or permuted configurations.

Molecular life is a great example of the sequence versatility. Comparative genomics provide the identification of numerous biologically functional elements. The nucleotide sequence of many genes, for example, is relatively well conserved between different species. In contrast, the sequences that regulate the gene expression are shorter and weaker. Thus, the simultaneous activation of a set of genes only can be explained in terms of conservation between configurations of higher-order regulatory elements, that can not be detected at the sequence level. We, therefore, have trained our meta-alignment programs in several datasets of regulatory regions collected from the literature. Then, we have tested the accuracy of our approximation to successfully characterize the promoter regions of human genes and their orthologs in other species. **Palabras Claves:** polinomios tipo Apostol generalizados; relación de recurrencia; ecuación diferencial; polinomios de Apostol-Bernoulli;

polinomios de Genocchi; polinomios de Jacobi; polinomios de Hermite; polinomios de Laguerre; polinomios de Charlier; polinomios de Bessel; polinomios de Bernoulli generalizados; números de Stirling de segunda clase.

En este trabajo estudiamos una nueva clase de polinomios tipo Apostol generalizados $Q_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$ con $(\alpha, u, v \in \mathbb{C}$ y $b, c \in \mathbb{R}^+)$ los cuales están definidos en un entorno adecuado de $t = 0$ por la siguiente función generatriz:

$$\left(\frac{(2^u t^v)^m}{\lambda b^t + \sum_{l=0}^{m-1} \frac{(t \log b)^l}{l!}} \right)^\alpha c^{xt} = \sum_{n=0}^{\infty} Q_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v) \frac{t^n}{n!}; \quad |t \log b| < |\log(-\lambda)|, \quad (7)$$

estos polinomios generan a las nuevas clases de todos los polinomios conocidos de Apostol, así:

$$\mathcal{B}_n^{[m-1, \alpha]}(x; b, c; \lambda) = (-1)^\alpha Q_n^{[m-1, \alpha]}(x; b, c; -\lambda; 0, 1), \quad (8)$$

$$\mathcal{E}_n^{[m-1, \alpha]}(x; b, c; \lambda) = Q_n^{[m-1, \alpha]}(x; b, c; \lambda; 1, 0), \quad (9)$$

$$\mathcal{G}_n^{[m-1, \alpha]}(x; b, c; \lambda) = Q_n^{[m-1, \alpha]}(x; b, c; \lambda; 1, 1). \quad (10)$$

Establecemos algunas propiedades básicas para estos polinomios, incluyendo relación de recurrencia y la ecuación diferencial que estos polinomios satisfacen. Finalmente se determinan fórmulas de conexión entre los polinomios y los polinomios de Genocchi, los polinomios de Jacobi, los polinomios de Hermite, los polinomios de Laguerre, los polinomios de Charlier, los polinomios de Bessel, los polinomios de Bernoulli generalizados $B_n^{[m-1]}(x)$ y los números de Stirling de segunda clase, lo cual extiende algunos resultados conocidos. Finalmente introducimos una nueva clase de polinomios tipo Apostol generalizados basados en los polinomios de Hermite de dos variables ${}_H Q_n^{[m-1, \alpha]}(x, b, c; \lambda; u, v)$ y mencionamos propiedades básicas.

Palabras Claves: polinomios tipo Apostol generalizados; relación de recurrencia; ecuación diferencial; poli-

nomios de Apostol-Bernoulli; polinomios de Genocchi; polinomios de Jacobi; polinomios de Hermite; polinomios de Laguerre; polinomios de Charlier; polinomios de Bessel; polinomios de Bernoulli generalizados; números de Stirling de segunda clase.

PARTE I

Preliminares

Preliminares

En el área de diseños de experimentos existen, principalmente, dos enfoques competitivos: uno basado en el análisis combinatorio más ajustado para modelos estadísticos de análisis de varianza y el otro enfoque basado en métodos analíticos que envuelve el análisis convexo, el cual se aplica, por ejemplo, a superficies de respuesta. Una concisa introducción a este último enfoque es dado en la breve monografía por Silvey (1980).

Mientras para modelos en la presencia de efectos fijos se han encontrado diseños óptimos para una gran variedad de casos (Schwabe, 2003) si los efectos de bloques se asumen que provienen de procesos aleatorios, aparecen dificultades adicionales. Un primer simple ejemplo de polinomios de regresión en la presencia de efectos de factores aleatorios discretos ha sido considerado en los pioneros artículos de Cheng (1995) y Atkins y Cheng (1999). Goos (2000) extendió este resultado a varias estructuras de bloque y obtuvo soluciones numéricas. A continuación presentaremos los conceptos básicos de la teoría de diseños de experimentos.

1.1 La ecuación de regresión

La siguiente ecuación es básica en la teoría de regresión:

$$Y_j = \eta(x_j, \theta) + \epsilon_j, \quad j = 1, \dots, N, \quad (1.1)$$

donde Y_1, \dots, Y_N son resultados experimentales, $\eta(x, \theta)$ es una función dada con vector de parámetros desconocidos $\theta = (\theta_1, \dots, \theta_m)^\top$, $\epsilon_1, \dots, \epsilon_N$ corresponden al error de observación y x_1, \dots, x_N son condiciones experimentales, que pertenecen a un conjunto compacto \mathcal{X} usualmente llamado la región de diseño.

Los casos para representar los resultados de los experimentos reales en la forma (1.1) se ha demostrado en muchos ejemplos, ver Rao (1973), Federow (1972), y Pukelsheim (1993).

Recordemos algunos supuestos básicos del modelo clásico de regresión.

- (a) Insesgamiento: $E(\epsilon_j) = 0$; ($j = 1, \dots, N$). Esto significa que $E(Y_j) = \eta(x_j, \theta)$ (es decir, el modelo está libre de un error sistemático).
- (b) Incorrelación: $E(\epsilon_i \epsilon_j) = 0$; ($i \neq j$).
- (c) Homogeneidad de varianza: $E(\epsilon_j^2) \equiv \sigma^2 > 0$; ($j = 1, \dots, N$).
- (d) La linealidad de parametrización: $\eta(x, \theta) = f(x)^\top \theta$, donde $f(x) = (f_1(x), \dots, f_m(x))^\top$, $f_i(x)$, $i = 1, \dots, m$, son funciones básicas y conocidas.

Como es habitual en la teoría estadística, estos supuestos proporcionan resultados observables a obtener y corresponden en cierta medida a las características de experimentos reales.

El propósito principal de un experimento es estimar un vector de parámetros desconocidos, o probar una hipótesis sobre los valores de los parámetros. Aquí, la exactitud de conclusiones estadísticas depende tanto del método de la inferencia estadística y en la elección de las condiciones experimentales.

Si (a) - (b) se asumen, entonces la técnica de mínimos cuadrados, proporciona valoración del vector θ de parámetros bajo cualesquiera condiciones experimentales fijas.

1.2 Estimación de mínimos cuadrados

La paternidad de este método se reparte entre Legendre que lo publicó en 1805 y Gauss que lo utilizó en 1795 y lo publicó en 1809.

El método de mínimos cuadrados (MC) es utilizado para estimar los parámetros en el modelo de regresión lineal.

Por ejemplo, en el modelo de regresión lineal múltiple

$$\begin{aligned} Y_j &= \beta_0 + \beta_1 x_{j1} + \cdots + \beta_k x_{jk} + \epsilon_j \\ &= \beta_0 + \sum_{i=1}^k \beta_i x_{ji} + \epsilon_j, j = 1, \dots, N. \end{aligned}$$

Suponga que se tiene $N > k$ observaciones. Se asume que $E(\epsilon_j) = 0$ y $\text{Var}(\epsilon_j) = \sigma^2$ y que los errores son independientes. El método de mínimos cuadrados minimiza la suma de cuadrados del error dada por

$$\begin{aligned} \text{SSE} &= \sum_{j=1}^N \epsilon_j^2 \\ &= \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right)^2 \end{aligned}$$

con respecto a cada uno de los parámetros del modelo $\beta_0, \beta_1, \dots, \beta_k$.

La derivada con respecto a β_0

$$\begin{aligned}\frac{\partial SSE}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0} \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right)^2 \\ &= -2 \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right)\end{aligned}$$

La derivada con respecto a β_j ($j = 1, 2, \dots, k$) es

$$\begin{aligned}\frac{\partial SSE}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right)^2 \\ &= -2 \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji}\end{aligned}$$

igualando a cero las derivadas, se tiene

$$\frac{\partial SSE}{\partial \beta_0} = -2 \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) = 0$$

$$\frac{\partial SSE}{\partial \beta_j} = -2 \sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} = 0$$

Simplificando para β_0 se tiene

$$\sum_{j=1}^N Y_j - \sum_{j=1}^N \beta_0 - \sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} = 0$$

$$\sum_{j=1}^N \beta_0 + \sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} = \sum_{j=1}^N Y_j$$

$$N\beta_0 + \sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} = \sum_{j=1}^N Y_j$$

$$N\beta_0 + \sum_{j=1}^N (\beta_1 x_{j1} + \beta_2 x_{j2} + \cdots + \beta_k x_{jk}) = \sum_{j=1}^N Y_j$$

$$N\beta_0 + \beta_1 \sum_{j=1}^N x_{j1} + \beta_2 \sum_{j=1}^N x_{j2} + \cdots + \beta_k \sum_{j=1}^N x_{jk} = \sum_{j=1}^N Y_j$$

Simplificando para β_j se tiene

$$\sum_{j=1}^N \left(Y_j - \beta_0 - \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} = 0$$

$$\left(\sum_{j=1}^N Y_j - \sum_{j=1}^N \beta_0 - \sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} = 0$$

$$\sum_{j=1}^N Y_j x_{ji} - \beta_0 \sum_{j=1}^N x_{ji} - \left(\sum_{j=1}^N \sum_{i=1}^k \beta_i x_{ji} \right) x_{ji} = 0$$

$$\sum_{j=1}^N Y_j x_{ji} - \beta_0 \sum_{j=1}^N x_{ji} - \sum_{j=1}^N (\beta_1 x_{j1} + \cdots + \beta_k x_{jk}) x_{ji} = 0$$

$$\sum_{j=1}^N Y_j x_{ji} - \beta_0 \sum_{j=1}^N x_{ji} - \left(\sum_{j=1}^N \beta_1 x_{j1} + \cdots + \beta_k \sum_{j=1}^N x_{jk} \right) x_{ji} = 0$$

$$\beta_0 \sum_{j=1}^N x_{ji} + \left(\sum_{j=1}^N \beta_1 x_{j1} + \cdots + \beta_k \sum_{j=1}^N x_{jk} \right) x_{ji} = \sum_{j=1}^N Y_j x_{ji}$$

Luego las ecuaciones normales son:

$$\begin{aligned}
 N\beta_0 + \beta_1 \sum_{j=1}^N x_{j1} + \cdots + \beta_k \sum_{j=1}^N x_{jk} &= \sum_{j=1}^N Y_j \\
 \beta_0 \sum_{j=1}^N x_{j1} + \beta_1 \sum_{j=1}^N x_{j1}^2 + \cdots + \beta_k \sum_{j=1}^N x_{jk}x_{j1} &= \sum_{j=1}^N Y_j x_{j1} \\
 \beta_0 \sum_{j=1}^N x_{j2} + \beta_1 \sum_{j=1}^N x_{j1}x_{j2} + \cdots + \beta_k \sum_{j=1}^N x_{jk}x_{j2} &= \sum_{j=1}^N Y_j x_{j2} \\
 &\vdots \\
 \beta_0 \sum_{j=1}^N x_{jk} + \beta_1 \sum_{j=1}^N x_{j1}x_{jk} + \cdots + \beta_k \sum_{j=1}^N x_{jk}^2 &= \sum_{j=1}^N Y_j x_{jk}
 \end{aligned}$$

Observe que hay $p = k + 1$ ecuaciones. Para obtener la solución es conveniente utilizar notación matricial. En esta notación el modelo se expresa como

$$Y = X\beta + \epsilon$$

donde

Y es el vector de observaciones

X es una matriz $n \times p$ de niveles de la variable

β es un vector $p \times 1$ de coeficientes de regresión

ϵ es el vector aleatorio error de orden $p \times 1$

La suma de cuadrados del error es dada por

$$SSE = \sum_{j=1}^N \epsilon_j^2 = \epsilon' \epsilon = (Y - X\beta)'(Y - X\beta)$$

Luego se obtiene que las ecuaciones normales son

$$X'X\hat{\beta} = X'Y$$

Para solucionar las ecuaciones normales se requiere que exista la inversa de la matriz $X'X$. Esta existe siempre que las variables regresoras sean linealmente independientes. Así, la solución de mínimos cuadrados de vector paramétrico β es

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Ejemplo 1

Myers y Montgomery (1985) describen un experimento que trata con motores de turbina de gas. El voltaje de salida de los motores se midió en diversas combinaciones de velocidad de la cuchilla y extensión del sensor de medición de tensión. Los datos del experimento se dan en el Cuadro 1.1.

El propósito del experimento fue estimar un modelo cuadrático completo con el voltaje de salida como la variable dependiente y la velocidad de la cuchilla y extensión del sensor de medición de tensión como las variables explicativas.

Variables codificadas

Es conveniente para la mayoría de aplicaciones describir el experimento en términos de variables

Corrida	Voltaje	Velocidad de la cuchilla (pulg/seg)	Extension (pulg)
1	1.23	5300	0.000
2	3.13	8300	0.000
3	1.22	5300	0.012
4	1.92	8300	0.012
5	2.02	6800	0.000
6	1.51	6800	0.012
7	1.32	5300	0.006
8	2.62	8300	0.006
9	1.65	6800	0.006
10	1.62	6800	0.006
11	1.59	6800	0.006

Tabla 1.1 Experimento turbina de gas descrito por Myers y Montgomery (1995).

codificadas, porque esto facilita la comparación de los diseños de diferentes experimentos. Por lo tanto, se reajustarán las variables cuantitativas. Es característico de una variable u cuantitativa o continua que varía entre un valor mínimo y máximo, u_{\min} y u_{\max} . Por lo general, los niveles de los factores se reajustarán a estar entre -1 y $+1$. Los valores codificados pueden ser calculados por

$$z = \frac{u - u_0}{\Delta}$$

donde u_0 es el punto medio del intervalo $[u_{\min}, u_{\max}]$ y Δ es la mitad de la diferencia entre u_{\max} y u_{\min} . Para la interpretación de los resultados experimentales, sin embargo, es deseable para volver a los niveles de los factores originales. Para el experimento de turbina de gas, los niveles codificados se pueden obtener de la siguiente manera:

$$x_1 = \frac{\text{Velocidad de la cuchilla} - 6800}{1500}$$

y

$$x_2 = \frac{\text{Extension} - 0.006}{0.006}$$

donde x_1 y x_2 representan los niveles codificados de los factores de velocidad de la cuchilla y extensión del sensor de medición de tensión, respectivamente. Los niveles codificados se muestran en el Cuadro 1.2. Vamos a utilizar esta forma de analizar los datos. Del Cuadro 1.2, es fácil ver que las corridas 9, 10 y 11 se llevan a cabo en el nivel medio de los factores experimentales.

Corrida	Voltaje	Velocidad de la cuchilla (pulg/seg)	Extension (pulg)
1	1.23	-1	-1
2	3.13	+1	-1
3	1.22	-1	+1
4	1.92	+1	+1
5	2.02	0	-1
6	1.51	0	+1
7	1.32	-1	0
8	2.62	+1	0
9	1.65	0	0
10	1.62	0	0
11	1.59	0	0

Tabla 1.2 Forma codificada del experimento de turbinas de gas.

Análisis

El propósito del experimento fue estimar un modelo cuadrático completo en las dos variables. Como resultado, la expansión polinómica

$$\mathbf{f}^T(\mathbf{x}) = \begin{bmatrix} 1 & x_1 & x_2 & x_1x_2 & x_1^2 & x_2^2 \end{bmatrix}$$

y

$$\boldsymbol{\beta}^T = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_{12} & \beta_{11} & \beta_{22} \end{bmatrix},$$

de manera que $p = 6$ y el modelo estadístico se pueden escribir como

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \epsilon.$$

La matriz de diseño de todo el experimento viene dada por

$$X = \begin{pmatrix} 1 & -1 & -1 & +1 & +1 & +1 \\ 1 & +1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 & +1 & +1 \\ 1 & 0 & -1 & 0 & 0 & +1 \\ 1 & 0 & +1 & 0 & 0 & +1 \\ 1 & -1 & 0 & 0 & +1 & 0 \\ 1 & +1 & 0 & 0 & +1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

la matriz $X^T X$ es

$$\begin{pmatrix} 11 & 0 & 0 & 0 & 6 & 6 \\ 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 6 & 0 & 0 & 0 & 6 & 4 \\ 6 & 0 & 0 & 0 & 4 & 6 \end{pmatrix}$$

y entonces $(X^T X)^{-1}$ es

$$\begin{pmatrix} 0.2632 & 0 & 0 & 0 & -0.1579 & -0.1579 \\ 0 & 0.1667 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 \\ -0.1579 & 0 & 0 & 0 & 0.3947 & -0.1053 \\ -0.1579 & 0 & 0 & 0 & -0.1053 & 0.3947 \end{pmatrix}$$

El vector Y es

$$Y = \begin{pmatrix} 1.23 \\ 3.13 \\ 1.22 \\ 1.92 \\ 2.02 \\ 1.51 \\ 1.32 \\ 2.62 \\ 1.65 \\ 1.62 \\ 1.59 \end{pmatrix}$$

y el vector $X^T Y$ es

$$\begin{pmatrix} 19.83 \\ 3.9 \\ -1.73 \\ -1.2 \\ 11.44 \\ 11.03 \end{pmatrix}$$

el estimador de mínimos cuadrados de β es

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\begin{aligned}
 \hat{\beta} &= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_{12} \\ \hat{\beta}_{11} \\ \hat{\beta}_{22} \end{pmatrix} \\
 &= \begin{pmatrix} 0.2632 & 0 & 0 & 0 & -0.1579 & -0.1579 \\ 0 & 0.1667 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 \\ -0.1579 & 0 & 0 & 0 & 0.3947 & -0.1053 \\ -0.1579 & 0 & 0 & 0 & -0.1053 & 0.3947 \end{pmatrix} \begin{pmatrix} 19.83 \\ 3.9 \\ -1.73 \\ -1.2 \\ 11.44 \\ 11.03 \end{pmatrix} \\
 &= \begin{pmatrix} 1.6705 \\ 0.6500 \\ -0.2883 \\ -0.3000 \\ 0.2237 \\ 0.0187 \end{pmatrix}
 \end{aligned}$$

luego el modelo ajustado por mínimos cuadrados es

$$\hat{Y} = 1.6706 + 0.6500x_1 - 0.2883x_2 - 0.3000x_1x_2 + 0.2237x_1^2 + 0.0187x_2^2$$

1.3 Propiedades de los estimadores de mínimos cuadrados

(1) $\hat{\beta}$ es un estimador insesgado de β . Ésto es, $E(\hat{\beta}) = \beta$.

Demostración:

$$\begin{aligned} E(\hat{\beta}) &= E[(X^T X)^{-1} X^T Y] \\ &= (X^T X)^{-1} X^T E(Y) \end{aligned}$$

y como $E(Y) = X\beta$, entonces

$$\begin{aligned} E(\hat{\beta}) &= (X^T X)^{-1} X^T X\beta \\ &= \beta \end{aligned}$$

(2) La matriz de varianzas y covarianzas del vector $\hat{\beta}$ es $\text{Cov}(\hat{\beta}) = \text{Var}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$

Demostración:

$$\text{Var}(\hat{\beta}) = \text{Var}((X^T X)^{-1} X^T Y)$$

Sea $A = (X^T X)^{-1} X^T$ y como A es una matriz y Y un vector columna y por la propiedad de varianzas $\text{Var}(AY) = A\text{Var}(Y)A^T$, se tiene que

$$\begin{aligned} \text{Var}(\hat{\beta}) &= (X^T X)^{-1} X^T \text{Var}(Y) (X^T X)^{-1} X^T \\ &= (X^T X)^{-1} X^T \sigma^2 ((X^T X)^{-1} X^T)^T \\ &= (X^T X)^{-1} X^T \sigma^2 (X^T)^T ((X^T X)^{-1})^T \\ &= (X^T X)^{-1} X^T \sigma^2 X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} (X^T X) (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} I \\ &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

Ejemplo 2

Para los datos del ejemplo 1, se tiene que la estimación de la matriz de varianzas-covarianzas del vector $\hat{\beta}$ es

$$\text{Var}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$$

$$= 0.0083 \begin{pmatrix} 0.2632 & 0 & 0 & 0 & -0.1579 & -0.1579 \\ 0 & 0.1667 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 \\ -0.1579 & 0 & 0 & 0 & 0.3947 & -0.1053 \\ -0.1579 & 0 & 0 & 0 & -0.1053 & 0.3947 \end{pmatrix}$$

$$= \begin{pmatrix} 0.00218456 & 0 & 0 & 0 & -0.00131057 & -0.00131057 \\ 0 & 0.00138361 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.00138361 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.002075 & 0 & 0 \\ -0.00131057 & 0 & 0 & 0 & 0.00327601 & -0.00087399 \\ -0.00131057 & 0 & 0 & 0 & -0.00087399 & 0.00327601 \end{pmatrix}$$

Los errores estándar de cada parámetro es dado en la tabla:

Parámetro	Error estándar
β_0	$\sqrt{0.00218456} = 0.04673928$
β_1, β_2	$\sqrt{0.00138361} = 0.03719691$
β_{12}	$\sqrt{0.002075} = 0.04555217$
β_{11}, β_{22}	$\sqrt{0.00327601} = 0.05723644$

1.4 Teorema de Gauss - Markov

Si no se asume normalidad el estimador mínimo cuadrático $\hat{\beta}$ es el mejor estimador dentro de los estimadores lineales insesgados de β , en el sentido que es el de la varianza más pequeña.

Demostración:

Conocemos que $\hat{\beta} = (X^T X)^{-1} X^T Y$, sea $A = (X^T X)^{-1} X^T$, entonces $\hat{\beta} = AY$.

Demostremos que $\beta = CY$ no es mejor estimador insesgado que $\hat{\beta}$. Sabemos que $E[CY] = CE[Y] = CX\beta$ por ser β insesgado se cumple que $E[\beta] = \beta \implies CX\beta = \beta \implies CX\beta - \beta = \vec{0} \implies (CX - I)\beta = \vec{0}$, como $\beta \neq 0$, entonces $\implies CX - I = \vec{0} \implies CX = I$.

Ahora como C es una matriz y Y un vector columna y por la propiedad $\text{Var}(AY) = A\text{Var}(Y)A^T$, tenemos que $\text{Var}(CY) = C\text{Var}(Y)C^T = C\sigma^2 C^T = \sigma^2 CC^T$.

Luego como $\text{Var}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ y $CX = I$ y $(CX)^T = X^T C^T = I$, podemos escribir $\text{Var}(\hat{\beta}) = CX\sigma^2(X^T X)^{-1}X^T C^T = \sigma^2 CX(X^T X)^{-1}X^T C^T$

por lo tanto

$$\begin{aligned} \text{Var}(\beta) - \text{Var}(\hat{\beta}) &= \sigma^2 CC^T - \sigma^2 CX(X^T X)^{-1}X^T C^T \\ &= \sigma^2 C(I - X(X^T X)^{-1}X^T)C^T \\ &= \sigma^2 CHC^T \end{aligned}$$

H es simétrica ($H^T = H$) e idempotente ($H^2 = H$), H es definida positiva (Una matriz A se dice que es semidefinida positiva si $Y^T AY \geq 0$ para todo vector $Y \neq 0$. Diremos que es definida positiva si $Y^T AY > 0$ para todo vector $Y \neq 0$), por tanto $CHC^T > 0$.

Entonces $\text{Var}(\beta) - \text{Var}(\hat{\beta}) > 0 \implies \text{Var}(\beta) > \text{Var}(\hat{\beta})$. Luego $\hat{\beta}$ es el mejor estimador insesgado de mínima varianza.

1.5 Diseños experimentales y matrices de información

El conjunto $\{x_1, \dots, x_N\}$ de elementos de \mathcal{X} (aunque algunos de los elementos pueden coincidir unos con otros) se llama un diseño exacto (o discreto) de tamaño N .

Tengamos sólo $n < N$ puntos distintos. Supongamos que x_i ocurre r_i veces entre los puntos

$\{x_1, \dots, x_N\}$ para $i = 1, \dots, n$ así

$$N = \sum_{i=1}^n r_i$$

.

Asociamos $w_i = r_i/N$ con cada uno de los puntos x_i , $i = 1, \dots, n$. Así

$$\xi = \begin{pmatrix} x_1 & \dots & x_n \\ w_1 & \dots & w_n \end{pmatrix} \quad (1.2)$$

será llamada diseño exacto (discreto) o diseño n -puntos de tamaño N .

La matriz

$$M(\xi) = \sum_{i=1}^n f(x_i) f^T(x_i) w_i \quad (1.3)$$

se llama la matriz de información de diseño ξ .

Por el teorema de Gauss-Markov, tenemos

$$\text{Cov}(\theta) = \frac{\sigma^2}{N} M^{-1}(\xi)$$

para la matriz varianza de la estimación de mínimos cuadrados.

El diseño ξ es una medida de probabilidad discreta, definido por (1.2), que incluye los puntos del conjunto \mathcal{X} y los coeficientes de peso.

En muchas situaciones prácticas, es imposible de realizar estos diseños y tales diseños se deben considerar como aproximación de algunos diseños discretos.

Vamos a escribir un diseño, concentrado en un número finito de puntos, en la forma (1.2), donde los coeficientes $w_i = \xi(x_i)$ son números positivos arbitrarios tales que $\sum w_i = 1$. La matriz

$$M(\xi) = \sum_{i=1}^n f(\mathbf{x}_i) f(\mathbf{x}_i)^\top \xi(\mathbf{x}_i) \quad (1.4)$$

se le llama la matriz de información del diseño aproximado.

Sea Ξ el conjunto de todos los diseños aproximados y \mathcal{M} el conjunto de matrices de información que les corresponden:

$$\mathcal{M} = \{M; M = M(\xi) \text{ para algunos } \xi \in \Xi\}$$

Sea Ξ_n el conjunto de diseños aproximados, concentrado en n puntos (con pesos distintos de cero).

Las propiedades básicas de matrices de información pueden enunciarse como un teorema.

TEOREMA 1.1. *Propiedades de las matrices de información:*

(i) *Cualquier matriz de información es definida no negativa (en particular es simétrica).*

Demostración:

La demostración fácilmente se concluye de la definición de matriz no negativa y del hecho de que para todo vector $z \in \mathbb{R}^m$ se cumple

$$z^\top M(\xi) z = \sum_{i=1}^n z^\top f(\mathbf{x}_i) f(\mathbf{x}_i)^\top z \xi(\mathbf{x}_i) = \sum_{i=1}^n \|z^\top f(\mathbf{x}_i)\|^2 \xi(\mathbf{x}_i) \geq 0$$

(ii) *Si $n < m$, siendo m el número de parámetros, entonces $\det M(\xi) = 0$.*

Demostración:

Supongamos que ξ tiene en su soporte $k < m$ puntos: $\mathbf{x}_1, \dots, \mathbf{x}_k$. En el desarrollo del determinante aparecerán siempre al menos dos columnas iguales y por tanto el determinante ha de ser cero.

(iii) El conjunto \mathcal{M} es convexo.

Demostración:

Es necesario comprobar que para cualquier $\lambda \in [0, 1]$ y para cualquier par de diseños ξ_1 y ξ_2 , la matriz

$$M = (1 - \lambda)M(\xi_1) + \lambda M(\xi_2)$$

pertenece al conjunto \mathcal{M} .

Definimos el diseño ξ , según la fórmula

$$\xi = (1 - \lambda)\xi_1 + \lambda\xi_2$$

mostraremos que $M = M(\xi) \in \mathcal{M}$; en efecto,

$$\begin{aligned} M &= (1 - \lambda)M(\xi_1) + \lambda M(\xi_2) \\ &= (1 - \lambda) \sum_{i=1}^n f(\mathbf{x}_i)f(\mathbf{x}_i)^\top \xi_1(\mathbf{x}_i) + \lambda \sum_{i=1}^n f(\mathbf{x}_i)f(\mathbf{x}_i)^\top \xi_2(\mathbf{x}_i) \\ &= \sum_{i=1}^n f(\mathbf{x}_i)f(\mathbf{x}_i)^\top [(1 - \lambda)\xi_1(\mathbf{x}_i) + \lambda\xi_2(\mathbf{x}_i)] \\ &= M(\xi) \end{aligned}$$

Con ello queda demostrado.

1.6 Criterios de optimalidad

Llamemos al diseño ξ no singular si el $\det M(\xi) \neq 0$. Vamos a considerar sólo el caso de la estimación de todo el conjunto de parámetros. Aquí, sólo los diseños no singulares son de interés. El teorema de Gauss-Markov es válido para ello.

Típicamente, no existe un diseño $\hat{\xi}$ tal que la matriz

$$M^{-1}(\hat{\xi}) - M^{-1}(\xi),$$

es no negativa, donde ξ es un diseño arbitrario. Por lo tanto, algunas funciones de matrices de información, que tienen sentido estadístico, se utilizan como los criterios de optimalidad.

Consideremos algunos criterios de optimalidad de nuestro interés.

1.6.1 D – criterio

El criterio de diseño más usado en las aplicaciones es el de D-optimalidad, en el que la varianza generalizada de las estimaciones de los parámetros, o su logaritmo se reduce al mínimo.

El diseño $\xi = \arg \max_{\xi} |M(\xi)| = \arg \min_{\xi} |(M(\xi))^{-1}|$ se llama diseño D-óptimo.

El funcional

$$\Psi[M(\xi)] = -\text{Log}|M(\xi)|$$

se llama criterio D-óptimo; donde el funcional Ψ definido en el espacio de elementos de la matriz de información de $M \in \mathcal{M}$, se le denomina criterio de optimización, ver Kiefer(1974).

D-optimalidad, criterio del determinante, equivale a minimizar el volumen del elipsoide de variación de los estimadores lineales e insesgados de los parámetros desconocidos.

1.6.2 G – criterio

Para diseños continuos la varianza normalizada de la respuesta predicha es

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

El criterio G-óptimo es de la forma

$$\max_{\mathbf{x} \in \mathcal{X}} d(\mathbf{x}, \xi) \rightarrow \inf_{\xi}.$$

Tenga en cuenta que para el diseño discreto normado ξ ,

$$d(\mathbf{x}, \xi) = \frac{\sigma^2}{N} \text{Cov}(\mathbf{f}(\mathbf{x})^\top \boldsymbol{\theta});$$

es decir $d(\mathbf{x}, \xi)$ es igual (a la constante de precisión) a la varianza de un valor, predicha por el modelo en el punto \mathbf{x} .

1.6.3 D_s – criterio

Son apropiados cuando el interés es estimar un subconjunto de s parámetros de todo el p – vector $\boldsymbol{\beta}$. Por lo tanto sin pérdida de generalidad, los términos del modelo se pueden dividir en dos grupos

$$E(Y) = \mathbf{f}^\top(\mathbf{x})\boldsymbol{\beta} = \mathbf{f}_1^\top(\mathbf{x})\boldsymbol{\beta}_1 + \mathbf{f}_2^\top \boldsymbol{\beta}_2$$

donde $\boldsymbol{\beta}_1 \in \mathbb{R}^s$ y $\boldsymbol{\beta}_2 \in \mathbb{R}^{p-s}$. Los $\boldsymbol{\beta}_1$ son los parámetros de interés y los $p - s$ parámetros $\boldsymbol{\beta}_2$ suelen ser tratados como parámetros molestia.

Un ejemplo es cuando $\boldsymbol{\beta}_1$ corresponde a los factores experimentales y $\boldsymbol{\beta}_2$ corresponde a los parámetros de las variables de bloqueo (Ver ejemplos, capítulo 15 en Optimum Experimental Designs, with SAS de A.C. Atkinson, A.N. Donev y R.D. Tobias). Un segundo ejemplo es cuando los experimentos están diseñados para comprobar la bondad de ajuste de un modelo.

Para obtener expresiones para el criterio de diseño y función de varianza relacionadas, dividimos la matriz de información como

$$\mathbf{M}(\xi) = \begin{pmatrix} \mathbf{M}_{11}(\xi) & \mathbf{M}_{12}(\xi) \\ \mathbf{M}_{12}^\top(\xi) & \mathbf{M}_{22}(\xi) \end{pmatrix}$$

La matriz de covarianza para la estimación de mínimos cuadrados de β_1 es $M^{11}(\xi)$, la submatriz superior izquierda $s \times s$ de $M^{-1}(\xi)$. Se puede verificar, a partir de los resultados de la inversa de una matriz particionada (por ejemplo, Fedorov 1972, p. 24), que

$$M^{11}(\xi) = \left\{ M_{11}(\xi) - M_{12}(\xi)M_{22}^{-1}(\xi)M_{12}^{\top}(\xi) \right\}^{-1}$$

En consecuencia, el diseño D_s -óptimo para β_1 maximiza el determinante

$$|M_{11}(\xi) - M_{12}(\xi)M_{22}^{-1}(\xi)M_{12}^{\top}(\xi)| = \frac{|M(\xi)|}{|M_{22}(\xi)|}.$$

Por otro lado, en la práctica se pueden utilizar teoremas que proporcionan herramientas para la construcción y el control de la optimización de un diseño. Si la atención se limita al parámetro β , se considera el conocido teorema general de equivalencia (cf. Kiefer y Wolfowitz(1990); cf. Silvey (1980); cf. Pukelsheim(2006)).

1.6.4 Teorema de equivalencia

El siguiente resultado de Kiefer y Wolfowitz (1960) es de gran importancia en la teoría del diseño experimental óptimo.

TEOREMA 1.2. (*Kiefer-Wolfowitz. Teorema de equivalencia*) *Para el modelo (1.1), existe un diseño D -óptimo bajo los supuestos clásicos de regresión y las siguientes condiciones son equivalentes:*

- (i) ξ^* es un diseño D -óptimo.
- (ii) ξ^* es un diseño G -óptimo.
- (iii) $\max_{x \in \mathcal{X}} d(x, \xi^*) = m$.

Por otra parte, todos los D -óptimos diseños tienen la misma matriz de información, y la función de predicción de la varianza $d(x, \xi^*)$ alcanza su máximo en los puntos de cualquier diseño D -óptimo con soporte finito.

Vale la pena subrayar que el teorema es cierto para los diseños que sean D-óptimo en la clase de diseño aproximado.

Este teorema no sólo establece la equivalencia entre D y G-criterios, sino que también da la importante condición necesaria y suficiente de D-optimalidad: Diseño ξ^* es D-óptimo si y sólo si $\max_{x \in \mathcal{X}} d(x, \xi^*) = m$.

la demostración del teorema se puede encontrar en Kiefer y Wolfowitz (1960). Muchos análogos del teorema de Kiefer-Wolfowitz se pueden encontrar en Kiefer (1974), y aparece en forma más general en Whittle (1973).

Ejemplo 3

Sea $\eta(x_j, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3$; un modelo polinomial con $x \in [-1, 1]$.

En el caso D-óptimo, se verificará a continuación que el diseño

$$\xi^* = \begin{pmatrix} -1 & -\sqrt{5}/5 & \sqrt{5}/5 & 1 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

es un diseño D-óptimo.

En efecto, bastará con mostrar que el diseño ξ^* verifica las condiciones del criterio de D-óptimalidad. Primero note que su matriz de momento es

$$\begin{aligned}
 M(\xi^*) &= \sum_{x \in \{-1, -\sqrt{5}/5, \sqrt{5}/5, 1\}} [1, x, x^2, x^3]^\top [1 \ x \ x^2 \ x^3] (1/4) \\
 &= \begin{pmatrix} 1 & 0 & 3/5 & 0 \\ 0 & 3/5 & 0 & 13/25 \\ 3/5 & 0 & 13/25 & 0 \\ 0 & 13/25 & 0 & 63/125 \end{pmatrix}
 \end{aligned}$$

Ahora

$$M^{-1}(\xi^*) = \begin{pmatrix} 13/4 & 0 & -15/4 & 0 \\ 0 & 63/4 & 0 & -65/4 \\ -15/4 & 0 & 25/4 & 0 \\ 0 & 65/4 & 0 & 75/4 \end{pmatrix}$$

Luego

$$d(x, \xi^*) = f^\top(x) M^{-1}(\xi^*) f(x)$$

$$= \frac{75}{4}x^6 - \frac{105}{4}x^4 + \frac{33}{4}x^2 + \frac{13}{4}$$

Por último

$\max d(x, \xi^*) = 4$ y dicha función tiene sus puntos críticos en los valores del diseño.

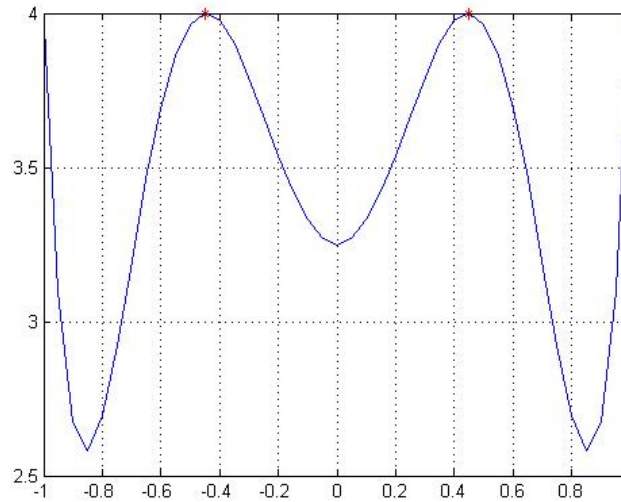


Figura 1.1 Ejemplo 3

Por lo tanto ξ^* es D-óptimo.

1.7 Diseño Factorial de Dos Factores Aleatorios

Suponga que se tienen dos factores, A y B, y que ambos tienen un gran número de niveles de interés. Se escogen al azar a niveles del factor A y b niveles del factor B. si el experimento se hace con n réplicas las observaciones pueden representarse con el modelo lineal:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

Donde todos los parámetros del modelo, $\tau_i, \beta_j, (\tau\beta)_{ij}, \epsilon_{ijk}$, son variables aleatorias independientes.

También se supondrá que las variables aleatorias $\tau_i, \beta_j, (\tau\beta)_{ij}, \epsilon_{ijk}$ siguen una distribución normal con media cero y varianzas $v(\tau_i) = \sigma^2$, $v(\beta_j) = \sigma^2$, $v[(\tau\beta)_{ij}] = \sigma^2$ y $v(\epsilon_{ijk}) = \sigma^2$ Por lo tanto.

$v(y_{ijk}) = \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2$ y $\sigma^2, \sigma^2, \sigma^2$ y σ^2 son los componentes de la varianza.

Las hipótesis que quieren probarse son $H_0 : \sigma^2_{\tau} = 0$ y $H_0 : \sigma^2_{\beta} = 0$

Los cálculos numéricos del análisis de varianza (SS_A, SS_B, SS_T, SS_E) se calculan como en el caso de efectos fijos. Para formar los estadísticos de prueba, deben examinarse los cuadrados medios

esperados, puede demostrarse que:

$$E(MS_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2$$

$$E(MS_B) = \sigma^2 + n\sigma_{\tau\beta}^2 + an\sigma_\beta^2$$

$$E(MS_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2$$

$$\text{y } E(MS_E) = \sigma^2$$

Por los cuadrados medios esperados se observa que el estadístico apropiado para probar la hipótesis de que no hay interacción, $H_0 : \sigma_{\tau\beta}^2 = 0$ es $F_0 = \frac{MS_{AB}}{MS_E}$

Ya que bajo H_0 tanto el numerador como el denominador de F_0 tienen valor esperado σ^2 , y solo si H_0 es falsa $E(MS_{AB})$ es mayor que $E(MS_E)$. El cociente F_0 se distribuye como $F_{(a-1), ab(n-1)}$ de manera similar para probar $H_0 : \sigma_\tau^2 = 0$ se usaría $F_0 = \frac{MS_A}{MS_{Ab}}$

Que se distribuye como $F_{a-1, (a-1)(b-1)}$ y para probar $H_0 : \sigma_\beta^2 = 0$ el estadístico es $F_0 = \frac{MS_B}{MS_{AB}}$

Que se distribuye como $F_{b-1, (a-1)(b-1)}$

Los componentes de la varianza pueden estimarse con el método del análisis de varianza

$$\hat{\sigma}^2 = MS_E$$

$$\hat{\sigma}_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n}$$

$$\hat{\sigma}_\beta^2 = \frac{MS_E - MS_{AB}}{an}$$

$$\hat{\sigma}_\tau^2 = \frac{MS_A - MS_{AB}}{bn}$$

PARTE II

Meta

Capítulo 2

Diseños óptimos en la presencia de efectos de bloques aleatorios

Para un modelo lineal en la presencia de efectos de bloques aleatorios se describe la situación donde s la j -ésima observación Y_{ij} al bloque i se puede escribir como

$$Y_{ij} = \gamma_i + \beta_0 + \mathbf{f}(\mathbf{x}_{ij})^\top \boldsymbol{\beta} + \epsilon_{ij} \quad (2.1)$$

donde \mathbf{x}_{ij} son los puntos experimentales, $j = 1, \dots, m_i$, $\mathbf{f} = (f_1, \dots, f_p)^\top$ es un conjunto de funciones (conocidas) de regresión y $\beta_0 \in \mathbb{R}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ son los parámetros desconocidos. f puede ser la función identidad para regresión lineal simple.

El término γ_i es el efecto del i -ésimo bloque aleatorio con $E(\gamma_i) = 0$ y $\text{Var}(\gamma_i) = \sigma_\gamma^2$. Los errores de observación aleatorio ϵ_{ij} se supone que son homoscedasticos, $E(\epsilon_{ij}) = 0$, $\text{Var}(\epsilon_{ij}) = \sigma^2$ y $\text{Cov}(\gamma_i, \epsilon_{ij}) = 0$. El análisis adicional dependerá del cociente de varianza $d = \sigma_\gamma^2 / \sigma^2$. Nos centraremos en los parámetros de la población $\theta = (\beta_0, \boldsymbol{\beta}^\top)^\top$. Asumiremos que el número de observaciones por bloque es constante, es decir, $m_i = m$.

Denotemos por $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^\top$ el vector de observaciones para el bloque i . La matriz de covarianza correspondiente $\text{Cov}(\mathbf{Y}_i) = \sigma^2 \mathbf{V}$ es completamente simétrica, $\mathbf{V} = \mathbf{I}_m + d \mathbf{1}_m \mathbf{1}_m^\top$, donde \mathbf{I}_m indica la matriz identidad $m \times m$ y $\mathbf{1}_m$ es un vector de longitud m con todas las entradas iguales a uno. El efecto fijo individual de la matriz de diseño $\mathbf{X}_i = (\mathbf{1}_m \mid \mathbf{F}_i)$ se puede descomponer en la primera columna de unos correspondiente a la intersección β_0 y la matriz de diseño para el vector de

parámetros β .

La inversa de V la podemos hallar mediante álgebra matricial

$$\begin{aligned}
 V^{-1} &= (I_m + d\mathbf{1}_m\mathbf{1}_m^\top)^{-1} \\
 &= I - d^{\frac{1}{2}}\mathbf{1}_m(I + d\mathbf{1}_m^\top I \mathbf{1}_m)^{-1}d^{\frac{1}{2}}\mathbf{1}_m^\top \\
 &= I - d\mathbf{1}_m(I + dmI)^{-1}\mathbf{1}_m^\top \\
 &= I - d\mathbf{1}_m(I(1 + dm))^{-1}\mathbf{1}_m^\top \\
 &= I - \frac{d}{1 + dm}\mathbf{1}_m\mathbf{1}_m^\top
 \end{aligned} \tag{2.2}$$

Luego, la matriz de información por bloque, utilizando (2.2) queda $\mathbf{X}_i^\top V^{-1} \mathbf{X}_i = \mathbf{X}_i^\top \mathbf{X}_i - \frac{d}{1+md} \mathbf{X}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{X}_i$ que es proporcional a la inversa de la matriz de varianza-covarianza $\text{Cov}(\hat{\theta}_i)$ si \mathbf{X}_i es de rango completo. Así θ es estimado sobre una base por bloque

$$\begin{aligned}
 \hat{\theta}_i &= (\mathbf{X}_i^\top V^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i^\top V^{-1} \mathbf{Y}_i \\
 &= (\mathbf{X}_i^\top \mathbf{X}_i)^{-1} \mathbf{X}_i^\top \mathbf{Y}_i
 \end{aligned}$$

Sobre la base de la población el mejor estimador lineal insesgado se puede calcular como $\hat{\theta} = \left(\sum_{i=1}^b \mathbf{X}_i^\top V^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^b \mathbf{X}_i^\top V^{-1} \mathbf{X}_i \hat{\theta}_i$ si d es conocido. Entonces $\text{Cov}(\hat{\theta}) = \sigma^2 \mathbf{M}_d^{-1}$, donde $\mathbf{M}_d = \sum_{i=1}^b \mathbf{X}_i^\top V^{-1} \mathbf{X}_i$ es la matriz de información sobre la base de la población. El subíndice d indica

la dependencia del cociente de varianzas d . Como $\mathbf{M}_d = \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{X}_i - \frac{d}{1 + md} \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{X}_i$.

La matriz de información particionada de acuerdo a β_0 y β , es

$$\begin{aligned}
\mathbf{M}_d &= \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{X}_i - \frac{d}{1+md} \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{X}_i \\
&= \sum_{i=1}^b (\mathbf{1}_m \mid \mathbf{F}_i)^\top (\mathbf{1}_m \mid \mathbf{F}_i) - \frac{d}{1+md} \sum_{i=1}^b (\mathbf{1}_m \mid \mathbf{F}_i)^\top \mathbf{1}_m \mathbf{1}_m^\top (\mathbf{1}_m \mid \mathbf{F}_i) \\
&= \sum_{i=1}^b \begin{pmatrix} \mathbf{1}_m^\top \\ \mathbf{F}_i^\top \end{pmatrix} (\mathbf{1}_m \mid \mathbf{F}_i) - \frac{d}{1+md} \sum_{i=1}^b \begin{pmatrix} \mathbf{1}_m^\top \\ \mathbf{F}_i^\top \end{pmatrix} \mathbf{1}_m \mathbf{1}_m^\top (\mathbf{1}_m \mid \mathbf{F}_i) \\
&= \sum_{i=1}^b \begin{pmatrix} m & \mathbf{1}_m^\top \mathbf{F}_i \\ \mathbf{F}_i^\top \mathbf{1}_m & \mathbf{F}_i^\top \mathbf{F}_i \end{pmatrix} - \frac{d}{1+md} \sum_{i=1}^b \begin{pmatrix} m & \\ \mathbf{F}_i^\top \mathbf{1}_m & \end{pmatrix} (m \mid \mathbf{1}_m^\top \mathbf{F}_i) \\
&= \sum_{i=1}^b \begin{pmatrix} m & \mathbf{1}_m^\top \mathbf{F}_i \\ \mathbf{F}_i^\top \mathbf{1}_m & \mathbf{F}_i^\top \mathbf{F}_i \end{pmatrix} - \frac{d}{1+md} \sum_{i=1}^b \begin{pmatrix} m^2 & m \mathbf{1}_m^\top \mathbf{F}_i \\ \mathbf{F}_i^\top \mathbf{1}_m m & \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \end{pmatrix} \\
&= \frac{1}{1+md} \sum_{i=1}^b \begin{pmatrix} m + m^2 d & \mathbf{1}_m^\top \mathbf{F}_i + md \mathbf{1}_m^\top \mathbf{F}_i \\ \mathbf{F}_i^\top \mathbf{1}_m + md \mathbf{F}_i^\top \mathbf{1}_m & \mathbf{F}_i^\top \mathbf{F}_i + md \mathbf{F}_i^\top \mathbf{F}_i \end{pmatrix} \\
&\quad - \frac{1}{1+md} \sum_{i=1}^b \begin{pmatrix} dm^2 & dm \mathbf{1}_m^\top \mathbf{F}_i \\ d \mathbf{F}_i^\top \mathbf{1}_m m & d \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \end{pmatrix} \\
\mathbf{M}_d &= \frac{1}{1+md} \left(\begin{array}{c|c} bm & \sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \\ \hline \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m & (1+md) \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - d \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \end{array} \right) \tag{2.3}
\end{aligned}$$

Si el interés está en los efectos fijos β solamente, entonces las reglas para invertir las correspondientes matrices de información parcial particionadas $\mathbf{M}_{\beta,d}^{-1} = \text{Cov}(\hat{\beta})/\sigma^2$ es igual a

$$\mathbf{M}_{\beta,d} = \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - \frac{d}{1+md} \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i - \frac{1}{bm} \frac{1}{1+md} \left(\sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \right) \left(\sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \right) \quad (2.4)$$

También consideramos los modelos límites para $d = 0$ y $d \rightarrow \infty$, respectivamente: Para $d = 0$ obtenemos los modelos de efectos fijos y sin interceptos de bloques

$$Y_{ij} = \beta_0 + \mathbf{f}(x_{ij})^\top \beta + \epsilon_{ij} \quad (2.5)$$

Obviamente, \mathbf{M}_d tiende a $\mathbf{M}_0 = \sum_{i=1}^b \mathbf{X}_i^\top \mathbf{X}_i$ para $d \rightarrow 0$. Del mismo modo, $\mathbf{M}_{\beta,d}$ tiende a

$$\mathbf{M}_{\beta,0} = \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - \frac{1}{bm} \left(\sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \right) \left(\sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \right) \quad (2.6)$$

Para $d \rightarrow \infty$ introducimos el modelo de efectos fijos con bloques fijos

$$Y_{ij} = \mu_i + \mathbf{f}(x_{ij})^\top \beta + \epsilon_{ij}; \quad (\mu_i = \gamma_i + \beta_0) \quad (2.7)$$

Aquí, el vector de parámetros $(\mu_1, \dots, \mu_b, \beta_1, \dots, \beta_p)^\top$ tiene dimensión $b + p$ y la matriz de información correspondiente tiene la forma

$$\mathbf{M}_{\infty} = \left(\begin{array}{ccc|c} & & & \mathbf{1}_m^{\top} \mathbf{F}_1 \\ & & & \vdots \\ & & & \mathbf{1}_m^{\top} \mathbf{F}_b \\ \hline & m\mathbf{I}_b & & \\ \hline \mathbf{F}_1^{\top} & \cdots & \mathbf{F}_b^{\top} \mathbf{1}_m & \sum_{i=1}^b \mathbf{F}_i^{\top} \mathbf{F}_i \end{array} \right) \quad (2.8)$$

Para β la matriz de información parcial correspondiente se puede calcular

$$\mathbf{M}_{\beta, \infty} = \sum_{i=1}^b \mathbf{F}_i^{\top} \mathbf{F}_i - \frac{1}{m} \sum_{i=1}^b \mathbf{F}_i^{\top} \mathbf{1}_m \mathbf{1}_m^{\top} \mathbf{F}_i \quad (2.9)$$

De ahí se obtiene el siguiente resultado, que establece la matriz de información parcial $\mathbf{M}_{\beta, d}$.

Lema1.

$$\mathbf{M}_{\beta, d} = \frac{1}{1 + md} \mathbf{M}_{\beta, 0} + \frac{md}{1 + md} \mathbf{M}_{\beta, \infty} \quad (2.10)$$

Tenga en cuenta que la matriz de información parcial $\mathbf{M}_{\beta, d}$ tiende a $\mathbf{M}_{\beta, \infty}$ cuando d tiende a ∞ .

2.1 Aspectos del diseño

La calidad de los estimadores $\hat{\theta}$ y $\hat{\beta}$ depende de la configuración experimental x_{ij} , $i = 1, \dots, b$, $j = 1, \dots, m$, a través de las matrices de información \mathbf{M}_d y $\mathbf{M}_{\beta, d}$, respectivamente. El objetivo en el diseño experimental es elegir los puntos de una región diseño \mathcal{X} con el fin de minimizar la covarianza $\text{Cov}(\hat{\theta})$ o $\text{Cov}(\hat{\beta})$ o partes de ella, lo cual es equivalente a maximizar las correspondientes matrices de información \mathbf{M}_d o $\mathbf{M}_{\beta, d}$ respectivamente. Como esas matrices no están completamente ordenadas, una optimización uniforme no es posible, en general. Por lo tanto, algunos funcionales de valores reales que ponen énfasis en las propiedades particulares de los estimadores se optimizarán. El criterio

de diseño más usado es el D-criterio, que tiene como objetivo maximizar el determinante de la matriz de información \mathbf{M}_d . Esto es equivalente a minimizar el volumen de un elipsoide de confianza para θ bajo la suposición de normalidad.

Si el interés está en los efectos β solamente, D_β -optimalidad se define en términos de la determinante de la inversa $\mathbf{M}_{\beta,d}^{-1}$ de la correspondiente matriz de información parcial. Como se ve a continuación,

$$\begin{aligned}
 \det(\mathbf{M}_d) &= \left| \frac{1}{1+md} \left(\begin{array}{c|c} bm & \sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \\ \hline \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m & (1+md) \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - d \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \end{array} \right) \right| \\
 &= \left(\frac{1}{1+md} \right)^{p+1} |bm| (1+md) \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - d \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \\
 &\quad - \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m (bm)^{-1} \sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i| \\
 &= \left(\frac{1}{1+md} \right)^{p+1} |bm| \left| \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i + md \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - d \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \right. \\
 &\quad \left. - \frac{1}{bm} \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \right| \\
 &= \left(\frac{1}{1+md} \right)^{p+1} |bm| \left| \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - \frac{1}{bm} \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \sum_{i=1}^b \mathbf{1}_m^\top \mathbf{F}_i \right. \\
 &\quad \left. + md \left(\sum_{i=1}^b \mathbf{F}_i^\top \mathbf{F}_i - \frac{1}{m} \sum_{i=1}^b \mathbf{F}_i^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_i \right) \right| \\
 &= \left(\frac{1}{1+md} \right)^{p+1} |bm| |\mathbf{M}_{\beta,0} + md \mathbf{M}_{\beta,\infty}| \\
 &= \left(\frac{1}{1+md} \right)^{p+1} |bm| (1+md) \left(\frac{1}{1+md} \mathbf{M}_{\beta,0} + \frac{md}{1+md} \mathbf{M}_{\beta,\infty} \right) | \\
 &= \left(\frac{1}{1+md} \right)^{p+1} |bm| (1+md) \mathbf{M}_{\beta,d} |
 \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{1+md} \right)^{p+1} |bm| (1+md)^p |\mathbf{M}_{\beta,d}| \\
&= \frac{bm}{1+md} \det(\mathbf{M}_{\beta,d})
\end{aligned}$$

sujeta por la fórmula para el determinante de matrices particionadas. Por lo tanto, D y D_β —optimalidad coinciden también en modelos de interceptos aleatorios, un hecho bien conocido en el ajuste de efectos fijos.

Lema2. *Un diseño (x_{ij}) es D —óptimo si y sólo si es D_β —óptimo.*

Si tenemos en cuenta los diseños que son uniformes en todos los bloques, es decir, en que los parámetros experimentales son los mismos para cada bloque, $x_{ij} \equiv x_j$, entonces la situación se simplifica radicalmente. En este caso las matrices de diseños individuales coinciden, $\mathbf{F}_i = \mathbf{F}_1$ y $\mathbf{X}_i = \mathbf{X}_1$, respectivamente, y \mathbf{X}_1 tiene que ser de rango columna completa para permitir estimabilidad de θ . Además, $\hat{\theta} = \frac{1}{b} \sum_{i=1}^b \hat{\theta}_i$ se reduce a la media de los valores ajustados de forma individual para los parámetros.

La matriz de covarianza estandarizada \mathbf{M}_d^{-1} se descompone de forma aditiva en la matriz correspondiente \mathbf{M}_0^{-1} para el modelo de efectos fijos y sin intercepciones individuales y la variabilidad de la intersección aleatoria (véase, por ejemplo, Entholzner y otros., (2005)). Para la matriz de información reducida observamos

$$\mathbf{M}_{\beta,0} = b \left(\mathbf{F}_1^\top \mathbf{F}_1 - \frac{1}{m} \mathbf{F}_1^\top \mathbf{1}_m \mathbf{1}_m^\top \mathbf{F}_1 \right) = \mathbf{M}_{\beta,\infty} \quad (2.11)$$

y, en consecuencia, por el Lema 1 $\mathbf{M}_{\beta,d} = \mathbf{M}_{\beta,0}$ es independiente de d . Así, el diseño D —óptimo para el modelo de efectos fijos y sin intersecciones individuales es D — y D_β —óptimo para cada $d \geq 0$ visto en el Lema 2.

Ejemplo de aplicación del modelo propuesto

Consideremos el modelo de regresión cuadrático en dos variables sin interacciones

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \epsilon_i; (x_{1i}, x_{2i}) \in [-1, 1] \times [-1, 1].$$

El diseño ξ el cual asigna iguales pesos $\frac{1}{9}$ a las cuatro esquinas $(\pm 1, \pm 1)$, a los cuatro puntos centrales de los lados $(0, \pm 1)$; $(\pm 1, 0)$ y al punto central $(0, 0)$ de la región experimental. Este diseño es D-óptimo para este modelo. Si el diseño ξ es bloqueado como sigue

$$\xi_1 = \begin{pmatrix} (-1, 0) & (0, 1) & (1, -1) \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} (-1, -1) & (0, 0) & (1, 1) \\ 1/3 & 1/3 & 1/3 \end{pmatrix},$$

$$\xi_3 = \begin{pmatrix} (-1, 1) & (1, 0) & (0, -1) \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

Entonces por el lema 2, ξ es D-óptimo para el correspondiente modelo en la presencia de efectos de bloques con respuesta

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_{11} x_{1ij}^2 + \beta_{22} x_{2ij}^2 + \gamma_i + \epsilon_{ij},$$

en la j -ésima corrida sobre el bloque i , ($i = 1, 2, 3$; $j = 1, 2, 3$) con pesos dados por $\xi(i, (x_{1ij}, x_{2ij})) = \frac{1}{3} \xi_i(x_{1ij}, x_{2ij})$. Además este diseño D-óptimo por bloques no depende del cociente de varianza d .

Conclusiones y trabajos futuros

En el presente trabajo se desarrolla la matriz de información de los criterios fijos en un modelo de regresión lineal en la presencia de efectos de bloques aleatorios como una combinación convexa de las matrices de información de los modelos límites cuando la varianza del efecto de bloque es cero o tiende a infinito.

Por otro lado, se muestra que los diseños óptimos para modelos de efectos fijos también son óptimos para modelos en la presencia de efectos de bloques aleatorios siempre que los bloques sean uniformes.

D y D_β -optimalidad coinciden también en modelos en la presencia de efectos de bloques aleatorios, un hecho ya conocido en los escenarios con efectos fijos.

Para trabajos futuros se pueden considerar modelos en la presencia de efectos de bloques aleatorios donde los parámetros de regresión interactúan con diferentes grupos o tratamientos.

- [1] Atkins and Cheng. (1998). Optimal regression designs in the presence of random block effects. JSPI 77, 321-335.
- [2] Cheng. Optimal regression designs under random blocks-effects models. (1995). Statistica Sinica 5, 485-497.
- [3] Debusho and Haines. (2007). V- and D-optimal population designs for the simple linear regression model with a random intercept term. Journal of Statistical Planning and inference 138, 1116-1130.
- [4] Entholzner, M., Benda, N., Schwabe, R. (2005). A note on designs for estimating population parameters. Biometrical Letters 42, 25-41.
- [5] Fedorov, V.V. (1972) Theory of Optimal Experiments. Academic Press, New York.
- [6] Fedorov, V.V. (1997). Model-Oriented Design of Experiments. Springer, New York.
- [7] Goos, Peter. (2002) The Optimal Design of Blocked and Split-Plot. Springer, New York.
- [8] GraBhoff, U and Schwabe, R. (2003). On the analysis of paired observations. Statistics and Probability Letters 65, 7-12.
- [9] Norell, L. (2006). Optimal designs for maximum likelihood estimators in the one-way random model. U.U.D.M. Report 2006:24. Department of Mathematics, Uppsala University.

- [10] Schmelter, T. and Schwabe, R. (2008). On optimal designs in random intercept models. *Tetra Mountains Mathematical Publications*. 39, 145-53.
- [11] Schwabe, R. (1996). *Optimum Designs for Multi-Factor Models*. Springer, New York.
- [12] Silvey, S.D. (1980). *Optimal Designs*. Chapman & Hall, London.
- [13] Van Breukelen, G.J.P., Candel, M.J.J.M. and Berger, M.P.F. (2008). Relative efficiency of unequal cluster sizes for variance component estimation in cluster randomized and multicentre trials. *Statistical Methods in Medical Research*, 17, 439-58.

PARTE III

Appendices

Ejemplo de Tesis

Al interior de los experimentos estadísticos la teoría de los diseños óptimos ha sido desarrollada. En general el tema de esta teoría es que para un apropiado modelo, si queremos poner énfasis sobre una cualidad particular de los parámetros a estimar, entonces la configuración experimental debería ser elegida de acuerdo a ciertos criterios con sentido estadístico. En la literatura relacionada con los diseños óptimos, un prominente autor fue Kiefer (1959), el cuál presentó los principales conceptos, tales como diseños aproximados y una variedad de criterios de optimalidad para esta rama de los diseños de experimentos; Kiefer, en particular dio el nombre D—optimalidad al criterio introducido por Wald (1943), este criterio es el más comunmente aplicado y está definido en función del determinante de la matriz de covarianza.

Más recientemente son reconocidos los libros de Atkinson y Donev (1992) y Pukelsheim (1993), donde los autores hacen una presentación estadística formal de los diseños óptimos. El presente trabajo se ha organizado en tres capítulos: el capítulo uno (Preliminares) contiene conceptos generales que sirven de apoyo y base a la teoría que se desarrolla en los siguientes dos capítulos. El capítulo dos trata sobre diseños óptimos en la presencia de efectos de bloques aleatorios, y en el capítulo tres se daran a conocer las conclusiones y una serie de problemas abiertos para futuras investigaciones relacionadas con el tema central de este trabajo de investigación.

Programa de Matemáticas

Universidad del Atlántico