

Una introducción a los diseños óptimos

An Introduction to Optimal Designs

VÍCTOR IGNACIO LÓPEZ^{1,2,a}, ROGELIO RAMOS^{2,b}

¹ESCUELA DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DE COLOMBIA, MEDELLÍN

²CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS (CIMAT), GUANAJUATO, GTO., MÉXICO

Resumen

Introducimos varios conceptos utilizados en la teoría de diseños de experimentos óptimos. Definimos criterios de optimalidad utilizados en esta área y exploramos sus propiedades. Se listan algunos resultados importantes para encontrar diseños óptimos para modelos lineales y no lineales, entre ellos teoremas de equivalencia. Finalmente se presentan algunos ejemplos típicos donde se aplica la teoría vista anteriormente.

Palabras clave: función de información, matriz de información, criterios de optimalidad, teoremas de equivalencia, modelos de regresión no lineal.

Abstract

We introduce several concepts used in optimal experimental design. Optimality criteria used in this area are defined and their properties are explored. Some important results for finding optimal designs in linear and nonlinear models are listed, specially equivalence theorems are formulated. Finally, we present some examples where that theory is applied.

Key words: Information function, Information matrix, Optimality criteria, Equivalence theorems, Nonlinear regression models.

1. Introducción

En muchas áreas de investigación interesa explicar una variable respuesta, Y , a través de k -variables explicativas, $x^T = [x_1, x_2, \dots, x_k]$, mediante un modelo de la forma:

$$Y = Y(x) = \eta(x, \theta) + \epsilon \quad (1)$$

^aProfesor asistente. E-mail: vilopez@unalmed.edu.co. Estudiante de doctorado en Ciencias con Orientación en Probabilidad y Estadística.

^bInvestigador titular. E-mail: rramosq@cimat.mx

siendo $\eta(x, \theta)$ una funci3n lineal o no lineal en el vector de parámetros desconocido $\theta \in \mathbb{R}^m$; y el término de error se asume que tiene media cero y varianza constante σ^2 . Una vez se especifica el modelo, la siguiente etapa consiste en determinar en qué condiciones experimentales, niveles de los x_j 's, se debe medir la respuesta para obtener una mejoría en la calidad de la inferencia estadística a un menor costo. Esto se logra construyendo un diseño donde la elecci3n de los niveles de los x_j 's y la frecuencia de medici3n de la respuesta están regidas por alg3n criterio de optimalidad (con significado estadístico). Hay varios ejemplos prácticos que han hecho uso de los diseños óptimos (véase Atkinson (1996)) y existe un gran número de contribuciones sobre este tema; por ejemplo, entre otros autores, Smith (1918) encontr3 diseños para los modelos polinomiales, Kiefer (1959) introdujo explícitamente la noci3n de diseño óptimo y sus propiedades; y posteriormente realiz3 muchos trabajos en el área (véase Brown et al. (1985)). Tambi3n, recientemente en los libros de Atkinson & Donev, A. N. (1992) y Pukelsheim (1993), los autores hicieron un tratamiento estadístico y formal, respectivamente, de los diseños óptimos.

Este trabajo tiene como objetivo presentar los conceptos b́sicos de los diseños óptimos y, en forma general, los criterios de optimalidad, tanto en modelos lineales como no lineales, dando mayor énfasis y extensi3n a los primeros, ya que son una alternativa de soluci3n para los modelos no lineales, por ejemplo los diseños óptimos locales mencionados en la secci3n 3.1.

Este art́culo se divide en cuatro secciones. En la siguiente secci3n se darán los aspectos sobresalientes de los diseños óptimos para el modelo lineal, se definen los criterios de optimalidad en general y se mencionan varios resultados, principalmente teoremas de equivalencia para determinar optimalidad. En la tercera secci3n se estudia el caso no lineal y se definen algunos de los criterios de optimalidad usados en la literatura. En la última secci3n se construyen diseños óptimos para dos posibles escenarios: cuando el experimentador conoce de antemano los puntos de soporte del diseño, caso usual en diseños de experimentos (véase la secci3n 4.1); y cuando no se conocen ni los puntos de soporte ni los pesos del diseño (véase la secci3n 4.3).

2. Diseños óptimos para modelos lineales

Para los modelos lineales se considera que la relaci3n entre las N -observaciones Y_i y x_i est́ dada por:

$$Y(x_i) = \theta^T f(x_i) + \epsilon, \quad x_i \in \mathbb{R}^k, \quad \theta \in \mathbb{R}^m$$

donde $f = [f_1, \dots, f_m]^T$ es un vector de m -funciones continuas linealmente independientes definidas en un conjunto compacto χ , rango de regresi3n, $\chi \subseteq \mathbb{R}^k$, $\theta \in \mathbb{R}^m$ es un vector de m -parámetros desconocidos, ϵ es una variable aleatoria con media cero y varianza constante σ^2 y se asume incorrelaci3n en las N -observaciones.

Aunado al modelo anterior, se define un diseño aproximado,

$$\xi = \begin{bmatrix} x_1 & \dots & x_n \\ w_1 & \dots & w_n \end{bmatrix}$$

con $w_i = \xi(x_i)$, como una medida de probabilidad definida en \mathcal{B} , conjunto de Borel de χ que incluye los conjuntos unitarios; tal que ξ tiene soporte finito. El soporte de ξ es $\text{Supp}(\xi) = [x_1, \dots, x_n]$, n : número de puntos de soporte de ξ , y las observaciones $Y(x)$ se hacen en x_1, \dots, x_n con frecuencias (o pesos) aproximadamente proporcionales a w_1, \dots, w_n .

Para cada diseño ξ se define la matriz de momentos:

$$M(\xi) \equiv \int_{\chi} f(x)f^T(x)d\xi(x) = \sum_{i=1}^n f(x_i)f^T(x_i)w_i$$

La forma de cuantificar la información suministrada por la matriz de momentos depende de los criterios de optimalidad, definidos como aquellos que maximizan algún funcional real (con un significado estadístico) de la matriz de momentos sobre Ξ ; clase de todos los diseños aproximados definidos en \mathcal{B} . Estos criterios de optimalidad se presentan a continuación, siguiendo el enfoque de Pukelsheim (1993), quien introduce la matriz de información C_K , función puente que da cuenta de la “información” contenida en combinaciones lineales de θ ; C_K es una función del conjunto de las matrices definidas no negativas de orden m , $NND(m)$, en el conjunto de las matrices simétricas de orden q , $Sim(q)$. Concluyendo con la noción de función de información ϕ . La matriz de información intuitivamente mide la información que aporta el sistema de parámetros $K^T\theta$, mientras que la función de información la cuantifica por medio de un número real.

En las observaciones 1 y 2 se presenta lo anterior esquemáticamente, y en la observación 3 se da la formulación del problema de diseño.

Observación 1. Se considera el caso general, cuando el investigador está interesado en la estimación de q -combinaciones lineales de θ . Es decir, la estimación del subsistema $K^T\theta$, donde $K \in \mathbb{R}^{m \times q}$ es conocida y $r(K) = q$.

- Sea ξ un diseño factible para $K^T\theta$, es decir, $\mathcal{C}(K) \subseteq \mathcal{C}(M(\xi))$, $\mathcal{C}(A)$ es el espacio generado por las columnas de la matriz A . Se define la matriz de información como la función:

$$C_K : NND(m) \rightarrow Sim(q)$$

tal que: $C_K(M(\xi)) = (K^T M^{-1} K)^{-1}$, A^{-} denota una inversa generalizada de A .

Por notación, $A \geq 0$ si y sólo si $A \in NND(m)$; $A \geq B$ si y sólo si $A - B \geq 0$.

La matriz de información es homogéneamente positiva ($C_K(\delta A) = \delta C_K(A)$, $A \geq 0$, $\delta > 0$), superaditiva ($C_K(A + B) \geq C_K(A) + C_K(B)$, $A, B \geq 0$), $Rango(C_K) \subseteq NND(q)$, cóncava ($C_K((1 - \alpha)A + \alpha B) \geq (1 - \alpha)C_K(A) + \alpha C_K(B)$, $A, B \in NND(m)$, $0 < \alpha < 1$) e isotónica ($A \geq B \Rightarrow C_K(A) \geq C_K(B)$).

- Si $K = I_m$, interesa estimar θ , y $M(\xi)$ es no singular, entonces $C_I(M(\xi)) = M(\xi)$. Es decir, la matriz de informaci3n coincide con la matriz de momentos; por esta raz3n en la literatura M tambi3n se llama matriz de informaci3n.

Observaci3n 2. Cuantificaci3n de la informaci3n suministrada para cada dise1o, ya sea por la matriz de momentos o la matriz de informaci3n, es definida a partir de una funci3n de valor real ϕ .

Sea ϕ un funcional de valor real, $\phi : NND(q) \rightarrow \mathbb{R}$. ϕ es una funci3n de informaci3n si es: homog3neamente positiva ($\phi(\delta C) = \delta\phi(C)$, $\delta > 0, C \geq 0$), superaditiva: $\phi(C + D) \geq \phi(C) + \phi(D)$, no negativa: ($\phi(C) \geq 0, C \geq 0$) y semicontinua superiormente (los conjuntos de nivel $\{\phi \geq \alpha\} = \{C \in NND(q) : \phi(C) \geq \alpha\}$ son cerrados para todo $\alpha \in \mathbb{R}$).

Para lo que sigue ϕ , denotar3 una funci3n de informaci3n.

Observaci3n 3. Formulaci3n del problema de dise1o.

- El problema de dise1o para el sistema parametral $K^T\theta$ consiste en encontrar un dise1o ξ^* que sea factible y que maximice, sobre todos los dise1os ξ factibles para $K^T\theta$, la funci3n de informaci3n:

$$\phi(C_K(M(\xi))) = \phi((K^T M(\xi)^- K)^{-1})$$

Por las propiedades de ϕ y C_K , principalmente la semicontinuidad superior y la compacidad de χ , el m3ximo anterior se alcanza para alg3n dise1o ξ .

- c -optimalidad. Si $K = c$, $c \in \mathbb{R}^{m \times 1}$ entonces el criterio asociado se denomina c -optimalidad; se puede mostrar que la 3nica funci3n de informaci3n es la identidad: $\phi(\delta) = \delta$ y el problema de dise1o se reduce a encontrar un dise1o ξ^* que sea factible para $c^T\theta$ y maximice la funci3n de informaci3n:

$$\phi(C_c(M(\xi))) = C_c(M(\xi)) = (c^T M(\xi)^- c)^{-1}$$

observe que el lado derecho representa el inverso de la varianza asociada al estimador 3ptimo para $c^T\theta$; luego los dise1os c -3ptimos son aquellos que minimizan la varianza de $c^T\hat{\theta}$.

A continuaci3n se exhibe una clase de funciones de informaci3n, denominada matriz de medias (*matrix means*), la cual contiene los criterios de optimalidad de mayor popularidad.

Sea $C \in NND(q)$, para $C > 0$:

$$\phi_p(C) = \begin{cases} \lambda_{\max}(C), & p = \infty; \\ \left[\frac{1}{q} \text{tr}(C^p) \right]^{1/p}, & p \neq 0, p \neq \pm\infty; \\ (\det(C))^{1/q}, & p = 0; \\ \lambda_{\min}(C), & p = -\infty. \end{cases} \quad (2)$$

Observación 4. Anotaciones sobre los criterios ϕ_p -óptimos.

- ϕ_p es función de información para $p \in [-\infty, 1]$.
- Si un diseño ξ maximiza el criterio anterior, se dice que el diseño es ϕ_p -óptimo ($p \in [-\infty, 1]$).
- Si $C = C_K(M(\xi)) = (K^T M(\xi) - K)^{-1}$ y $p \in \{0, -1, -\infty\}$ se tienen los criterios de optimalidad más populares, versión generalizada, que dependen de la maximización del respectivo funcional evaluado en la matriz información (o en algunos casos evaluado en la matriz de diseño); ellos son, respectivamente,
 - D -optimalidad, criterio del determinante, equivale a minimizar el volumen del elipsoide asociado a la estimación del sistema $K^T \theta$, cuando los errores son normales.
 - A -optimalidad, criterio promedio, recíproco del promedio de las varianzas asociado a las q -combinaciones lineales de θ , y
 - E -optimalidad, criterio del valor propio, minimización del valor propio más pequeño.

El problema de optimización planteado en la observación 3 es muy complejo; en la práctica se hace uso de teoremas de equivalencia para verificar si un diseño dado es ϕ -óptimo (Pukelsheim 1993, Atkinson & Donev, A. N. 1992). El primer teorema de equivalencia lo demostraron Kiefer & Wolfowitz (1960); allí establecieron la equivalencia entre D -optimalidad y G -optimalidad – ξ es un diseño G -óptimo si minimiza: $\forall \xi \in \Xi$,

$$\bar{d}(M(\xi)) = \begin{cases} \sup_{x \in \chi} d(x, M(\xi)), & \mathcal{C}(M(\xi)) \supseteq \chi; \\ \infty, & \text{en otro caso.} \end{cases}$$

siendo, $d(x, M(\xi)) = f^T(x)M(\xi)^{-1}f(x)$. Es decir, si ξ minimiza la varianza más grande posible sobre χ , rango de regresión.

Teorema 1. *Teorema de Equivalencia de Kiefer-Wolfowitz.*

Sea $\chi \subseteq \mathbb{R}^k$ con m -vectores linealmente independientes. Un diseño ξ con matriz de momentos $M(\xi)$, definida positiva, es D -óptimo si y sólo si ξ es G -óptimo si y sólo si $f^T(x)M(\xi)^{-1}f(x) \leq m, \forall x \in \chi$ si y sólo si $\bar{d}(M(\xi)) = m$.

En caso de optimalidad, $f^T(x_i)M^{-1}f(x_i) = m, \quad \xi(x_i) \leq \frac{1}{m}, \quad \forall x_i \in \text{Supp}(\xi)$.

Por lo popular de los criterios ϕ_p ($p \in [-\infty, 1]$), se enuncia el siguiente teorema de equivalencia, da condiciones necesarias y suficientes para garantizar que un diseño dado es ϕ_p -óptimo.

Teorema 2. *Sea $\phi_p, p \in (-\infty, 1]$, \mathcal{M} un subconjunto convexo y compacto de $NND(m)$ y $M(\xi) \in \mathcal{M}$, con ξ factible para $K^T \theta$ y matriz de información $C = C_K(M(\xi))$. Entonces:*

- ξ es ϕ_p -3ptimo para $K^T\theta$ en \mathcal{M} si: $\exists G \in M^-$ tal que:

$$\text{Tr}(AGKC^{p+1}K^TG^T) \leq \text{Tr}(C^p), \quad \forall A \in \mathcal{M} \quad (\text{desigualdad de normalidad}).$$

En caso de optimalidad, la igualdad se obtiene si en vez de A se coloca M u otra matriz $\widetilde{M} \in \mathcal{M}$ ϕ_p -3ptima para $K^T\theta$ en \mathcal{M} .

- Si $0 < M(\xi) \in \mathcal{M}$, entonces ξ es ϕ_p -3ptimo para θ en \mathcal{M} si: $\text{Tr}(AM^{p-1}) \leq \text{Tr}(M^p), \forall A \in \mathcal{M}$.

Para $p = 0$ y $M > 0$, la condici3n requerida se traduce en: $\text{Tr}(AM^{-1}) \leq m, \forall A \in \mathcal{M}$, pero \mathcal{M} es generado por las matrices de rango uno: $A = f(x)f^T(x)$; es suficiente verificar la condici3n para A , y el lado izquierdo de la desigualdad es:

$$\text{Tr}(AM^{-1}) = \text{Tr}(f(x)f^T(x)M^{-1}) = \text{Tr}(f^T(x)M^{-1}f(x)) = f^T(x)M^{-1}f(x)$$

lo cual muestra un caso particular de una de las equivalencias del Teorema 1. Existe la versi3n del teorema de equivalencia para E -optimalidad ($p = -\infty$) (v3ase Pukelsheim 1993). Para $p = -1$ (A -optimalidad), $M > 0$ y $C = (K^TM^{-1}K)^{-1}$, la condici3n a verificar ser3:

$$f^T(x)M^{-1}KK^TM^{-1}f(x) - \text{Tr}(K^TM^{-1}K) \leq 0, \quad \forall x \in \chi \quad (3)$$

3. Dise3os 3ptimos para los modelos no lineales

Los modelos no lineales se pueden representar por:

$$Y(x) = \eta(x, \theta) + \epsilon \quad (4)$$

donde, como en el modelo lineal, las variables explicativas $x^T = [x_1, x_2, \dots, x_k]$ varían en un espacio de dise3o compacto, $\chi \subseteq \mathbb{R}^k$, dotado de una σ -3lgebra, \mathcal{B} , (Borelianos en χ , agreg3ndole los conjuntos unitarios), $\theta \in \Theta \subseteq \mathbb{R}^m$, los errores con media cero y varianza constante y $\eta(x, \theta)$ es una funci3n no lineal en θ .

En el modelo 4, dado un dise3o ξ definido en \mathcal{B} , se sabe que el estimador de m3nimos cuadrados para θ , bajo ciertas condiciones de regularidad, es asint3ticamente insesgado y su matriz de varianzas-covarianzas asint3tica es la inversa de la matriz:

$$M(\xi, \theta) = E_\xi [g(x, \theta)g^T(x, \theta)] = \int_\chi g(x, \theta)g^T(x, \theta) d\xi(x)$$

donde: $g(x, \theta) = \frac{\partial \eta(x, \theta)}{\partial \theta}$. Lo cual motiva el an3lisis de $M(\xi, \theta)$. En la literatura a M se le conoce como matriz de informaci3n, y juega el papel de la matriz de momentos del modelo lineal, si se considerara el modelo linealizado.

La dependencia de M de θ hace que la b3squeda de dise3os 3ptimos dependa de este par3metro. En forma an3loga al caso lineal, se cuantifica la magnitud de la informaci3n suministrada por $M(\xi, \theta)$ a partir de funcionales de 3sta, y consecuentemente la maximizaci3n de alguna funci3n de informaci3n ϕ , de valor real. Para la construcci3n de los dise3os 3ptimos existen varios enfoques; en este trabajo se exploran los siguientes:

3.1. Diseños óptimos locales

Introducidos por Chernoff (1953), son los primeros diseños que aparecieron para el caso no lineal. Consisten en dar inicialmente un valor a priori para θ , θ_0 , que esté cercano al valor verdadero del parámetro, luego utilizar la aproximación lineal de Taylor para $\eta(x, \theta)$ alrededor de θ_0 y construir diseños óptimos para el modelo linealizado: $Y^*(x) = \beta^T g(x, \theta_0) + \epsilon^*$. Los diseños resultantes son diseños óptimos locales. Varios autores han construido diseños con este enfoque; véase por ejemplo: Ford et al. (1992), Dette et al. (2004), Dette et al. (2005), entre otros. La construcción de diseños D -óptimos locales y A -óptimos locales se explora en los ejemplos de la sección 4.3.

3.2. Diseños óptimos promediados por una distribución a priori π -enfoque Bayesiano

Este criterio hace uso del conocimiento que se tiene acerca de θ por una distribución a priori π , resultando un criterio de optimalidad denominado Bayesiano. En particular, un diseño ξ es D -óptimo Bayesiano (con respecto a la distribución a priori π), para abreviar D_π -óptimo, si maximiza:

$$E_\theta[\log |M(\xi, \theta)|] = \int_{\Theta} \log |M(\xi, \theta)| d\pi(\theta)$$

En general, un diseño es ϕ -óptimo Bayesiano con respecto a la distribución a priori π , abreviado por ϕ_π -óptimo, si maximiza: $E_\theta[\phi(M(\xi))]$ (Dette et al. 2003). Ejemplos de este tipo de diseños se muestran en la sección 4.3.

Para D_π optimalidad, se obtiene la siguiente equivalencia, generalización del teorema de Kiefer y Wolfowitz:

$$\xi \text{ es } D_\pi - \text{óptimo si } E[g^T(x, \theta)M^{-1}(\xi, \theta)g(x, \theta)] \leq m, \quad \forall x \in \chi \quad (5)$$

La respectiva equivalencia se obtiene para A_π -optimalidad al calcular la esperanza, con respecto a θ , de la expresión 3:

$$\xi \text{ es } A_\pi - \text{óptimo si } E[g^T(x, \theta)M^{-1}(\xi)KK^TM^{-1}g(x, \theta) - \text{Tr}(K^TM^{-1}(\xi)K)] \leq 0, \quad \forall x \in \chi \quad (6)$$

donde K , M^{-1} , son funciones que dependen de θ .

4. Ejemplos

En esta sección se presentan varios ejemplos de modelos (lineales y no lineales) donde el interés está en encontrar diseños óptimos, ya sea que se conozcan los puntos de soporte o no. Inicialmente se considera el caso lineal, criterios ϕ_p optimales y por último el caso no lineal.

4.1. Ejemplo 1. Determinación de los pesos óptimos para un diseño dado

Este ejemplo muestra los resultados reportados por Pukelsheim & Torsney (1991) cuando los puntos de soporte del diseño son conocidos, y luego se da una aplicación.

Por conveniencia se reescribe el modelo lineal de la siguiente forma:

$$E[Y_{ij}] = x_i^T \theta, \quad j = 1, 2, \dots, n_i, \quad i = 1, 2, \dots, l \quad (7)$$

con observaciones Y_{ij} incorrelacionadas, varianza constante σ^2 , y los l vectores de regresión $\{x_1, x_2, \dots, x_l\}$ linealmente independientes y *conocidos*.

El objetivo es encontrar un diseño experimental ξ que indique, en forma óptima, el número de réplicas n_i que se harán en el vector de regresión x_i , con el fin de estimar $K^T \theta$. En términos generales, hallar un vector de pesos $w^T = [w_1, w_2, \dots, w_l]$ que maximice la función de información:

$$\phi_p[C_K(M(w))]$$

donde $M(w)$, la matriz de momentos asociada al modelo (7), es expresada como:

$$M(w) = \sum_{i=1}^l x_i x_i^T w_i = X^T \Delta_w X$$

$X^T = [x_1 \quad x_2 \quad \dots \quad x_l]$ y $\Delta_w = \text{diag}(w)$, con la siguiente inversa generalizada para M :

$$M(w)^- = X^T (X X^T)^{-1} \Delta_w^- (X X^T)^{-1} X$$

$V = (X X^T)^{-1} X K$, entonces la matriz de información es:

$$C_K(M(w)) = (V^T \Delta_w^- V)^{-1}$$

Δ_w^- inversa generalizada para Δ_w , si todos los pesos de w no son positivos.

En el siguiente resultado se obtiene una expresión cerrada para los pesos A -óptimos ($p = -1$) y una forma de encontrarlos recursivamente para los otros valores de p .

Teorema 3. *Sea $p \in (-\infty, 1]$, el vector de pesos w es ϕ_p -óptimo para $K^T \theta$ si y sólo si:*

$$w_i = \frac{\sqrt{b_{ii}}}{\sum_{j=1}^l \sqrt{b_{jj}}}, \quad \text{para } i = 1, \dots, l \quad (8)$$

donde b_{11}, \dots, b_{ll} son los elementos de la diagonal de la matriz definida no negativa $l \times l$: $B = V C^{p+1} V^T$, con $C = C_K(M(w))$.

Observación 5. Si $p = -1$, A -óptimalidad, y el sistema de interés es el vector de parámetros θ , entonces la función objetivo: $\phi_{-1}(M(w)) = \left[\frac{1}{m} \text{Tr}(M^{-1}(w)) \right]^{-1}$, el inverso del promedio de las varianzas de los estimadores de mínimos cuadrados $\hat{\theta}_1, \dots, \hat{\theta}_m$, estandarizados relativo a su tamaño muestral N y a la varianza del modelo σ^2 .

Aplicación: Modelo de análisis de varianza de un factor con tres niveles:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad j = 1, 2, \dots, n_i, \quad i = 1, 2, 3$$

En este caso, según el modelo (7), $\theta^T = [\mu_1, \mu_2, \mu_3]$, $x_1^T = [1, 0, 0]$, $x_2^T = [0, 1, 0]$, $x_3^T = [0, 0, 1]$. El interés está en conocer el número de réplicas en cada nivel del factor con el fin de estimar en forma óptima: *C1*.— Los tres efectos promedio, y *C2*.— El contraste: $\mu_3 - \mu_1$ y μ_2 . Aplicando el teorema 3 se obtienen los pesos ϕ_p -óptimos para $p = -1, 0$, ver tabla 1. Note que:

- Para el caso *C1* los diseños óptimos coinciden para los dos criterios considerados.
- En ambos casos, el diseño *A*-óptimo requiere la misma proporción de observaciones en cada uno de los tres niveles del factor. Difiere con respecto al criterio *D*-óptimo ya que en el caso *C2*, el diseño *D*-óptimo requiere alrededor de la mitad de las observaciones para el segundo nivel, y el resto se reparte igualmente para los otros dos niveles.

TABLA 1: Resultados para los pesos óptimos diseño de un factor.

	p	criterio	w
Caso 1. Estimación de μ	-1	A-optimalidad	$[1/3 \quad 1/3 \quad 1/3]$
	0	D-optimalidad	$[1/3 \quad 1/3 \quad 1/3]$
Caso 2. Estimación de $\mu_3 - \mu_1$ y μ_2	-1	A-optimalidad	$[1/3 \quad 1/3 \quad 1/3]$
	0	D-optimalidad	$[0.251 \quad 0.498 \quad 0.251]$

4.2. Ejemplo 2. Diseños óptimos para modelos polinomiales

Considere inicialmente el modelo polinomial de grado 2 en el intervalo $[-1, 1]$,

$$Y(x) = f^T(x)\theta + \epsilon$$

donde $f^T(x) = [1 \quad x \quad x^2]$ y $\theta^T = [\theta_0 \quad \theta_1 \quad \theta_2]$, $x \in [-1, 1]$.

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

$$\xi = \begin{bmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

es un diseño *D*-óptimo para estimar θ (tomando $K = I$).

En efecto, bastará con mostrar que el diseño ξ verifica las condiciones del teorema 1. Primero note que su matriz de momentos es:

$$M(\xi) = \int_{x \in \text{Supp}(\xi)} f(x)f^T(x) d\xi(x) = \sum_{x \in \{-1, 0, 1\}} [1 \quad x \quad x^2]^T [1 \quad x \quad x^2] 1/3 = \begin{bmatrix} 3 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

y para $x \in [-1, 1]$,

$$d(x, \xi) = f^T(x)M^{-1}(\xi)f(x) = [1 \quad x \quad x^2] \left(\frac{1}{3} \begin{bmatrix} 3 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} = \frac{9}{2}x^4 - \frac{9}{2}x^2 + 3$$

En la figura 1 se muestra que esta funci3n tiene todos sus valores por debajo de $m = 3$, y en los puntos de soporte alcanza su ḿximo, luego ξ es D -3ptimo para estimar el vector de parámetros θ .

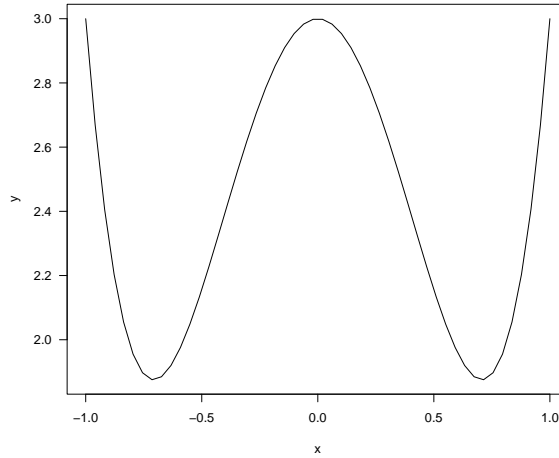


FIGURA 1: Gráfico de la funci3n $d(x, \xi)$.

Suponga que el inter3s del investigador est3 en estimar la diferencia entre el coeficiente de la potencia cuadr3tica y la lineal. En este caso el sistema de inter3s es: $K^T\theta = [0 \quad -1 \quad 1] \theta$.

En la tabla 2 aparecen los resultados que se obtuvieron con los criterios D y A optimalidad para los casos $C1$ y $C2$, y con los mismos puntos de soporte. Observe que los diseños dados por ambos criterios, para estimar $\theta_2 - \theta_1$, reparten en forma equitativa el n3mero de observaciones en los puntos $x = -1$ y en $x = 0$ y ninguna observaci3n para $x = 1$. Se presentan diferencias en los diseños 3ptimos para la estimaci3n del vector de parámetros; para A -optimalidad el 50% de las observaciones se deber3n tomar en $x = 0$, y el resto se reparte equitativamente en los otros dos puntos, mientras que con D -optimalidad el mismo n3mero de observaciones se deber3 tomar en los tres puntos.

En la literatura (Pukelsheim 1993) existe la soluci3n para el caso general, polinomios de grado d , para los diseños D -3ptimos en el intervalo $[-1, 1]$; los autores usan como argumento el teorema 1, y muestran que los diseños D -3ptimos tienen

igual peso $1/(d+1)$ en los puntos de soporte que son solución a la ecuación:

$$(1-x^2)\dot{P}_d(x) = 0$$

donde $\dot{P}_d(x)$ es la derivada del polinomio de Legendre de grado d .

TABLA 2: Resultados para los pesos óptimos para el modelo cuadrático.

	p	Criterio	w
Caso 1. Estimación de θ	-1	A-optimalidad	[0.25 0.50 0.25]
	0	D-optimalidad	[1/3 1/3 1/3]
Caso 2. Estimación de $\theta_2 - \theta_1$	-1	A-optimalidad	[0.5 0.5 0.0]
	0	D-optimalidad	[0.5 0.5 0.0]

4.3. Ejemplo 3. Modelos no lineales

Como ilustración se consideran dos modelos no lineales (Atkinson & Donev, A. N. 1992), y se construyen diseños óptimos locales y usando un enfoque Bayesiano.

1. El modelo de decaimiento exponencial está dado por:

$$\eta(x, \theta) = \exp(-\theta x), \quad x > 0$$

Si θ_0 es una buena asignación para θ , su matriz de información, la cual es un escalar, es: $M(\xi, \theta_0) = M(x_0, \theta_0) = \int_{x>0} f^2(x, \theta_0) d\xi(x)$, donde $f(x, \theta_0) = \frac{d}{d\theta}\eta(x, \theta)|_{\theta=\theta_0} = -x \exp(-\theta_0 x)$.

El modelo linealizado consta de un parámetro, y el diseño D -óptimo local concentra toda su masa en un punto. Se verá a continuación que el punto es: $x_0 = 1/\theta_0$. Sea ξ_0 el diseño que tiene como punto de soporte a x_0 , entonces:

$$M(\xi_0, \theta_0) = x_0^2 \exp(-2\theta_0 x_0) \quad (9)$$

No es difícil mostrar que el máximo de la ecuación (9) se alcanza en $x_0 = 1/\theta_0$, y

$$d(x, \xi_0) = f^T(x, \theta_0)M^{-1}(\xi_0, \theta_0)f(x, \theta_0) = \frac{f^2(x, \theta_0)}{\int f^2(x)d\xi_0(x)} = (x\theta_0)^2 \exp(-2(x\theta_0 - 1))$$

observe que $d(x, \xi_0) \leq 1, \forall x > 0$ y $d(x, \xi_0) = 1$ en $x = 1/\theta_0$, luego el diseño que concentra su masa en $1/\theta_0$ es D -óptimo local. Este diseño no permite realizar pruebas de bondad de ajuste para el modelo en cuestión. El diseño depende de la especificación de θ_0 , y puede llegar a ser ineficiente si θ_0 está muy lejos del valor verdadero θ . Otra forma de hallar un diseño óptimo es a partir de un enfoque Bayesiano, donde se incorpora el conocimiento acerca de θ por medio de una distribución a priori. Como ilustración se consideran 6 distribuciones a priori discretas, uniformes en 5 puntos, y se hallaron los

respectivos diseños D_π –óptimos para estimar θ , con las diferentes a priori; ver tabla 3. Lo anterior se hizo numéricamente con ayuda de algoritmos computacionales programados en el lenguaje *R* (R Development Core Team 2006) usando la equivalencia 5. Los diseños D_π –óptimos obtenidos están formados por tres puntos de soporte, observándose variación en las distintas a priori consideradas, tanto en los puntos de soporte como en sus pesos. Atkinson & Donev, A. N. (1992, pág. 230) muestran cómo los puntos de soporte del diseño aumentan a medida que la distribución a priori que ellos consideran es más dispersa.

TABLA 3: Ejemplo de decaimiento exponencial con diferentes distribuciones a priori uniformes para θ .

Puntos de la a priori j					x			w^*		
[0.09	0.49	1	4.9	9]	[0.156	1.503	10.998]	[0.438	0.403	0.158]
[0.10	0.50	1	5.0	10]	[0.143	1.517	9.812]	[0.432	0.420	0.148]
[0.11	0.51	1	5.1	11]	[0.132	1.536	8.812]	[0.428	0.437	0.135]
[0.12	0.52	1	5.2	12]	[0.123	1.558	7.952]	[0.424	0.455	0.121]
[0.14	0.54	1	5.4	14]	[0.107	1.617	6.547]	[0.418	0.496	0.085]
[0.15	0.55	1	5.5	15]	[0.101	1.649	5.965]	[0.416	0.521	0.063]

2. Modelo de compartimientos.

Los modelos de compartimientos son de gran utilidad en farmacocinética, utilizados, entre otras aplicaciones, para modelar el nivel de concentración de un medicamento en la sangre de un individuo a lo largo del tiempo. Se considera el siguiente modelo:

$$\eta(x, \theta) = \frac{\theta_1}{\theta_1 - \theta_2} \{ \exp(-\theta_2 x) - \exp(-\theta_1 x) \}, \quad x \geq 0, \quad \theta_1 > \theta_2 > 0 \quad (10)$$

Asociado al trabajo biológico es de interés, además de estimar el vector de parámetros θ , estimar tres cantidades que ayudan al estudio de la cinética del medicamento en un individuo. Estas cantidades son:

- a) El área bajo la curva (AUC): $g_1(\theta) = \int_0^\infty \eta(x, \theta) dx = \frac{1}{\theta_2}$.
- b) Tiempo para la concentración máxima: $g_2(\theta) = x_{\max} = \frac{\log \theta_1 - \log \theta_2}{\theta_1 - \theta_2}$.
- c) La concentración máxima: $g_3(\theta) = \eta(x_{\max}, \theta)$.

La construcción de diseños óptimos para la estimación de estas funciones simultáneamente, se hará por medio de diseños A –óptimos locales (véase (2), con $p = -1$), y diseños A –óptimos promediados por una distribución a priori uniforme. La j –ésima columna K_j , de K es el gradiente de función no lineal $g_j(\theta)$, evaluada en θ_0 . Así se asegura que el diseño óptimo será aquel que minimice el promedio de las varianzas del respectivo estimador linealizado, es decir, aquel que minimiza:

$$\frac{1}{3} \sum_{j=1}^3 \text{var}(g_j(\hat{\theta}_0)) \approx \frac{1}{3} \sum \text{var}(K_j^T \hat{\beta}) \propto K^T M^{-1}(\xi) K$$

En el caso de estudio, las primeras dos columnas de K están dadas por:

$$K_1^T(\theta_0) = [0 \quad -1/\theta_{20}^2], \quad K_2(\theta_0) = \left[\frac{1/\theta_{10} - x_{\max}}{\theta_{10} - \theta_{20}} \quad \frac{x_{\max} - 1/\theta_{20}}{\theta_{10} - \theta_{20}} \right]$$

en forma análoga se halla la tercera columna de K . Como ilustración, se tomó $\theta_0^T = [0.7 \quad 0.2]$, y en la tabla 4 se presentan los diseños A -óptimos locales obtenidos para la estimación de las tres características de interés simultáneamente. También se consideró una a priori uniforme discreta para los siguientes cinco valores del vector de parámetros θ :

$$\Theta = \{(0.70, 0.20), (0.65, 0.15), (0.75, 0.25), (0.65, 0.25), (0.75, 0.15)\}$$

es decir, $\pi(\theta) = 1/5$, $\forall \theta \in \Theta$, y en la tabla 4 se reporta el diseño A -óptimo Bayesiano obtenido. Ambos diseños A -óptimo local y A -óptimo promediado por la a priori π , presentan pocas diferencias. Además en ambos casos se verificó que el diseño hallado satisfacía las equivalencias dadas por (3) y (6), respectivamente.

TABLA 4: Diseños A -óptimos locales y promediados por la a priori π para el modelo 10.

Criterio	Tiempo x	Pesos del diseño
A -óptimalidad local	[1.313 6.602]	[0.276 0.724]
A -óptimalidad promediado por la a priori π	[1.456 7.145]	[0.269 0.731]

5. Anotaciones finales

En este trabajo se presentó una motivación inicial para el estudio de los diseños óptimos en ambos casos lineal y no lineal. Se dio el enfoque matemático de cada uno de los criterios de optimalidad usados en la práctica y se terminó presentando algunos ejemplos típicos. Hay gran diversidad de bibliografía en torno a este tema, donde el estudio en esta área es factible e interesante.

En la mayoría de los artículos citados, los autores asumen que el modelo bajo consideración es conocido, y el valor de los parámetros es desconocido. Con este supuesto, usan criterios de optimalidad que son eficientes para la estimación de los parámetros del modelo fijo. Sin embargo, existen aplicaciones donde la forma de la función de regresión no es conocida en forma exacta, es decir, el experimentador debe decidir, entre un conjunto de clases de funciones competitivas, cuáles de estas describen los datos en una forma más adecuada. Como lo afirman Biedermann et al. (2005), el problema de diseño para discriminar entre modelos no lineales competitivos ha encontrado muy poco interés en la literatura que aquellos problemas de estimación de parámetros. En el caso de discriminación de modelos, lineal

y no lineal, se pueden revisar los trabajos de: Atkinson & Cox (1974), Atkinson & Fedorov (1975), Pukelsheim & Rosenberger (1993), Biswas & Chaudhuri (2002) y Biedermann et al. (2005). Por lo anterior, est como trabajo futuro ahondar en el estudio de dise˜os ptimos que sean eficientes para discriminar entre modelos no lineales anidados, adems de que permitan estimar en forma simultnea funciones de los parmetros.

Agradecimientos

Agradecemos los comentarios hechos por los dos rbitros, lo que hizo que este trabajo se mejorara considerablemente. El presente trabajo se realiz cuando el primer autor estaba haciendo su doctorado en Ciencias con Orientacin en Probabilidad y Estadística en el Centro de Investigacin en Matemticas (CIMAT), Mxico. Parte de este trabajo fue apoyado por CIMAT, Secretara de Relaciones Exteriores de Mxico (SRE) y la Universidad Nacional de Colombia, sede Medelln.

Recibido: noviembre de 2006

Aceptado: marzo de 2007

Referencias

- Atkinson, A. C. (1996), ‘The Usefulness of Optimum Experimental Designs’, *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 59–76.
- Atkinson, A. C. & Cox, D. R. (1974), ‘Planning Experiments for Discriminating Between Models’, *Journal of the Royal Statistical Society. Series B (Methodological)* **36**(3), 321–348.
- Atkinson, A. C. & Donev, A. N. (1992), *Optimum Experimental Designs*, Oxford Science Publications, New York.
- Atkinson, A. C. & Fedorov, V. V. (1975), ‘Optimal Design: Experiments for Discriminating Between Several Models’, *Biometrika* **62**(2), 289–303.
- Biedermann, S., Dette, H. & Pepelyshev, A. (2005), ‘Optimal Discrimination Designs for Exponential Regression Models’, Preprint.
- Biswas, A. & Chaudhuri, P. (2002), ‘An Efficient Design for Model Discrimination on Parameter Estimation in Linear Models’, *Biometrika* **89**(3), 709–718.
- Brown, L. D., I., O., Sacks, J. & Wynn, H. P. (1985), *Jack Karl Kiefer Collected Papers III, Design of Experiments*, Springer Verlag, New York.
- Chernoff, H. (1953), ‘Locally Optimal Designs for Estimating Parameters’, *The Annals of Mathematical Statistics* **24**(24), 586–602.
- Dette, H., Haines, L. M. & Imhof, L. A. (2003), ‘Maximin and Bayesian Optimal Designs for Regression Models’, Preprint. pp. 1-15.

- Detle, H., Melas, V. B. & Pepelyshev, A. (2004), 'Optimal Designs for a Class of Nonlinear Regression Models', *The Annals of Statistics* **32**(5), 2142–2167.
- Detle, H., Melas, V. B. & Wong, W. K. (2005), 'Optimal Design for Goodness-of-Fit of the Michaelis-Menten Enzyme Kinetic Function', *Journal of the American Statistical Association* **100**(472), 1370–1381.
- Ford, I., Tornsney, B. & Wu, C. F. J. (1992), 'The Use of a Canonical Form in the Construction of Locally Optimal Designs for Nonlinear Problems', *Journal of the Royal Statistical Society, Series B (Methodological)* **54**.
- Kiefer, J. (1959), 'Optimum Experimental Designs (With Discussion)', *Journal Royal Statistical Society, B* **21**, 272–319.
- Kiefer, J. & Wolfowitz, J. (1960), 'The Equivalence of Two Extremum Problems', *Canadian Journal of Mathematics* **12**, 363–366.
- Pukelsheim, F. (1993), *Optimal Design of Experiments*, John Wiley & Sons, New York.
- Pukelsheim, F. & Rosenberger, J. L. (1993), 'Experimental Designs for Model Discrimination', *Journal of the American Statistical Association* **88**(422), 642–649.
- Pukelsheim, F. & Torsney, B. (1991), 'Optimal Weights for Experimental Designs on Linearly Independent Support Points', *The Annals of Statistics* **19**(3), 1614–1625.
- R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*<http://www.R-project.org>
- Smith, K. (1918), 'On the Standard Deviations of Adjusted and Interpolates Values of an Observed Polynomial Functions and its Constants and the Guidance They Give Towards a Proper Choice of the Distribution of Observations', *Biometrika* **12**, 1–85.