



CIMAT

Centro de Investigación en Matemáticas, A.C.

---

---

# Diseños óptimos para discriminación y estimación en modelos no lineales

T E S I S

Que para obtener el grado de  
**Doctor en Ciencias**

con Orientación en  
**Probabilidad y Estadística**

P r e s e n t a

**Víctor Ignacio López Ríos**

Director de Tesis:

**Dr. Rogelio Ramos Quiroga**

Guanajuato, Gto., México, Junio de 2008

# Centro de Investigación en Matemáticas, A.C.

Diseños óptimos para discriminación y estimación  
en modelos no lineales

TESIS  
que para obtener el grado de  
**Doctor en Ciencias**  
con Orientación en  
**Probabilidad y Estadística**  
presenta

**Víctor Ignacio López Ríos**

Integrantes del Jurado:

Dr. José Andrés Christen Gracia (CIMAT)  
Presidente

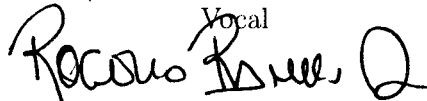
Dr. Jorge Domínguez Domínguez (CIMAT)  
Secretario

Dr. Jesús Fernando López Fidalgo (UNIVERSIDAD DE  
CASTILLA-LA MANCHA)

Vocal

Dr. Timothy O'Brien (LOYOLA UNIVERSITY OF CHICAGO)

Vocal



Dr. Rogelio Ramos Quiroga (CIMAT)

Vocal y Director de Tesis

(Vo. Bo. en representación del Jurado)



**CIMAT**  
CENTRO DE INVESTIGACION  
EN MATEMÁTICAS A. C.

# Centro de Investigación en Matemáticas, A.C.

## Acta de Examen de Grado

En la ciudad de Guanajuato, Gto., siendo las 16:00 horas del día 27 de junio del año 2008, se reunieron los miembros del jurado integrado por los señores:

Acta No.: 004

Libro No.: 003

Foja No.: 004

**DR. JOSÉ ANDRÉS CHRISTEN GRACIA** (CIMAT)  
**DR. JORGE DOMÍNGUEZ DOMÍNGUEZ** (CIMAT)  
**DR. ROGELIO RAMOS QUIROGA** (CIMAT)  
**DR. JESÚS FERNANDO LÓPEZ FIDALGO** (UNIV. DE CASTILLA-LA MANCHA)  
**DR. TIMOTHY O'BRIEN** (LOYOLA UNIVERSITY OF CHICAGO)

bajo la presidencia del primero y con carácter de secretario el segundo, para proceder a efectuar el examen que para obtener el grado de

**DOCTOR EN CIENCIAS  
CON ORIENTACIÓN EN PROBABILIDAD Y ESTADÍSTICA**

sustenta

**VÍCTOR IGNACIO LÓPEZ RÍOS**

en cumplimiento con lo establecido en los reglamentos y lineamientos de estudios de posgrado del Centro de Investigación en Matemáticas, A.C., mediante la presentación de la tesis

**"DISEÑOS ÓPTIMOS PARA DISCRIMINACIÓN Y ESTIMACIÓN EN MODELOS NO LINEALES".**

Los miembros del jurado examinaron alternadamente al (la) sustentante y después de deliberar entre sí resolvieron declararlo (a):

*Aprobado*

*[Signature]*  
 DR. JOSÉ ANDRÉS CHRISTEN GRACIA  
 Presidente

*[Signature]*  
 DR. JORGE DOMÍNGUEZ DOMÍNGUEZ  
 Secretario

*[Signature]*  
 DR. ROGELIO RAMOS QUIROGA  
 Vocal

*[Signature]*  
 DR. JESÚS FERNANDO LÓPEZ FIDALGO  
 Vocal

*[Signature]*  
 DR. TIMOTHY O'BRIEN  
 Vocal



**CIMAT**  
DIRECCION  
GENERAL

*[Signature]*  
 DR. JOSÉ CARLOS GÓMEZ LARRAÑAGA  
 Director General

# Dedicatoria

## A DIOS

Quién me lleno de fortaleza para culminar este proyecto de vida.

## A MI PADRE

**Víctor Daniel López Londoño** porque desde el lugar donde está sé que siempre estuvo conmigo.

## A MI MADRE

**Ana Lilliam Ríos** quién con su apoyo silencioso me dio ánimos y fuerzas para iniciar y terminar este trabajo en un hermoso y acogedor país, México.

## A LA FAMILIA LÓPEZ RÍOS

**Mi madre, hermanos y sobrinos** a quiénes les dedico esta nueva meta alcanzada en el camino de la vida.

# Agradecimientos

Al **Dr. Rogelio Ramos Quiroga** por su apoyo, paciencia y don de gentes que permitió que esta investigación llegara a un feliz término.

A la **Universidad Nacional de Colombia, sede Medellín**, en especial a la Escuela de Estadística, por el apoyo brindado para la realización y culminación de mis estudios doctorales, lo cual redundará en beneficio de la comunidad Universitaria.

A la **Secretaría de Relaciones Exteriores (SRE) de México** y al **CONACYT**, por la beca otorgada de agosto 2005 a julio 2007.

Al **CIMAT**, por los apoyos complementarios otorgados en becas y pagos de colegiatura para los periodos agosto 2004 a julio 2005 y agosto 2007 a junio 2008.

A los **proyectos CONACYT 45974 - F**, responsable Dra. Graciela González Farías y **47972**, responsable Dr. Johan Van Horebek; por los apoyos complementarios recibidos para la participación en congresos internacionales.

A los **investigadores del área de probabilidad y estadística** por sus enseñanzas y orientaciones.

A **mis compañeros y amigos**, Patricia Bautista, Oyuki Hermosillo, Antonio Muriello, Jair Remigio, José Montoya, Oscar Dalmau y Fidel Ulín, quiénes hicieron mi estancia en México amena, agradable e inolvidable.

# Contenido

<b>1. Introducción</b>	<b>1</b>
<b>2. Farmacocinética y modelos de compartimientos</b>	<b>5</b>
2.1. Introducción . . . . .	5
2.2. Conceptos de farmacocinética . . . . .	5
2.3. Modelos de Compartimientos . . . . .	7
2.3.1. Modelos de compartimientos y diagramas de sistemas . . . . .	8
2.3.2. Solución del sistema de ecuaciones diferenciales . . . . .	11
<b>3. Teoría de los diseños óptimos</b>	<b>15</b>
3.1. Introducción . . . . .	15
3.2. Diseños óptimos para modelos lineales . . . . .	16
3.3. Eficiencia de los diseños . . . . .	28
3.4. Teoremas de equivalencia . . . . .	29
3.5. Consecuencias del teorema general de equivalencia . . . . .	34
3.6. Diseños óptimos para modelos no lineales . . . . .	37
3.6.1. Diseños óptimos locales . . . . .	39
3.6.2. Diseños óptimos promediados por una distribución a priori $\pi$ . . . . .	40
3.6.3. Ejemplos de diseños óptimos en modelos no lineales . . . . .	41
3.7. Criterios de optimalidad para discriminar entre modelos . . . . .	44
3.7.1. $\mathbb{T}$ -optimalidad . . . . .	45
3.7.2. Algunos ejemplos de diseños $\mathbb{T}$ -óptimos . . . . .	46
3.8. Resumen Criterios de optimalidad . . . . .	52
<b>4. Diseños óptimos: Discriminación y estimación</b>	<b>53</b>
4.1. Introducción . . . . .	53
4.2. Metodología Propuesta . . . . .	54
4.3. Ejemplo ilustrativo de la propuesta . . . . .	61
4.3.1. Modelo de cuatro compartimientos . . . . .	62
4.3.2. Modelo de tres compartimientos . . . . .	70
4.3.3. Diseños óptimos promediados por una distribución a priori . . . . .	76

<i>CONTENIDO</i>	v
4.3.4. Diseños óptimos para discriminar entre ambos modelos . . . . .	85
4.3.5. Diseños $DEFNS_\alpha$ -óptimos locales . . . . .	87
4.3.6. Comparación entre los diferentes criterios de optimalidad locales	92
4.3.7. Diseños $DEFNS_\alpha^\pi$ -optimalidad Bayesiano . . . . .	95
<b>5. Consideraciones Adicionales</b>	<b>105</b>
5.1. Introducción . . . . .	105
5.2. Implicaciones prácticas . . . . .	105
5.3. Diseños óptimos exactos, bajo independencia . . . . .	107
5.4. Diseños óptimos exactos, con estructura de covarianza . . . . .	108
<b>6. Conclusiones</b>	<b>112</b>
<b>A.</b>	<b>116</b>
A.1. Algoritmo para $\phi$ -optimalidad . . . . .	116
A.2. Algoritmo para $\mathbb{T}$ -optimalidad . . . . .	117
A.3. Procedimiento alternativo . . . . .	117
<b>B.</b>	<b>120</b>
B.1. Gradiente de la función $\eta_C(t; \Theta)$ , modelo $I$ . . . . .	120
B.2. Características de interés, modelo $I$ . . . . .	120
B.3. Gradiente de la función $\eta_C(t; \mathbb{B})$ , modelo $II$ . . . . .	124
B.4. Características de interés, modelo $II$ . . . . .	125
<b>Bibliografía</b>	<b>128</b>
<b>Anexos</b>	<b>134</b>

# Capítulo 1

## Introducción

En muchas áreas donde se hace uso de la estadística, como en Biología, Medicina, Química, Ingeniería, entre otras, se necesitan hacer experimentos con el fin de obtener información acerca de un fenómeno bajo estudio. Los resultados obtenidos dependen en gran medida de la buena planeación que se hace del experimento, además del presupuesto y tiempo disponible para realizarlo. En la mayoría de los experimentos son de gran relevancia los costos y el tiempo para la obtención de la información. Es importante, entonces, contar con una metodología estadística que garantice en cierta medida la obtención de la información en determinados factores asociados al problema, factores elegidos no arbitrariamente sino a partir de algún criterio estadístico que mida las bondades del objetivo u objetivos que se busquen con la investigación. La teoría de los diseños óptimos resuelve el problema anterior desde el punto de vista de maximización de funcionales de la matriz de información, con alguna interpretación estadística. En otras palabras, en este contexto, un diseño para un modelo dado contiene las condiciones experimentales o los niveles de uno o varios factores y también las repeticiones en cada uno de los niveles del factor o de los factores.

Se tiene conocimiento de la construcción de estos diseños desde la segunda década del siglo pasado con el trabajo de Smith (1918), donde halla diseños óptimos en modelos polinomiales hasta de grado seis. Posteriormente este procedimiento fue denominado  $\mathbb{G}$ -optimización por Kiefer y Wolfowitz (1959). Otros trabajos son presentados en: Kiefer (1959), Hoel y Levine (1964), Karlin y Studden (1966), Pukelsheim y Torsney (1991), Dette, Melas y Pepelyshev (2004a), entre otros. Los libros de Atkinson y Donev (1992) y Pukelsheim (1993), representan un buen compendio de los resultados más importantes obtenidos hasta esos momentos.

En la literatura, el criterio denominado del determinante,  $\mathbb{D}$ -optimalidad, propuesto por Wald (1943), tiene como propósito encontrar las localizaciones de los puntos donde se tomarán las observaciones buscando maximizar el determinante de la matriz de in-



formación de Fisher. Así como este criterio existen una gran diversidad de variantes que son funcionales de esta matriz y serán el objeto de estudio del capítulo 3. Todos los criterios están relacionados de alguna forma con la varianza ya sea de los estimadores de los parámetros del modelo o del estimador de una o varias funciones de los parámetros. En el caso de modelos no lineales los diseños dependerán de los parámetros del modelo, de ahí que el diseño obtenido será efectivo en la medida de que tan confiable es la información acerca de los parámetros Box y Lucas (1959).

Así como existen criterios asociados con la estimación de los parámetros, también los hay en el caso donde se tiene como objetivo discriminar entre varios modelos competitivos que describen el mismo fenómeno, Atkinson y Cox (1974), Atkinson y Fedorov (1975a), Atkinson y Fedorov (1975b), Fedorov y Khabarov (1986), Felsenstein (1992), Muller y Ponce de Leon (1996), Ponce de Leon y Atkinson (1991), Ponce de Leon y Atkinson (1992), Waterhouse et al. (2008), Pukelsheim y Rosenberger (1993). Además, varios autores han puntualizado que los diseños obtenidos para discriminación, en general, no son eficientes para la estimación de los parámetros del modelo considerado, y más aun tampoco lo serán si en el problema de estudio se desean estimar algunas funciones de los parámetros, (Atkinson y Donev (1992)). En modelos lineales hay varios trabajos en esta dirección donde se buscan criterios que discriminen y estimen los parámetros del modelo simultáneamente, ver Spezzaferri (1988), Zen y Tsai (2002), Biswas y Chaudhuri (2002) y Pukelsheim y Rosenberger (1993), entre otros. En el caso no lineal se encuentran los trabajos de Borth (1975) quién propone un criterio a partir de la entropía total para el problema de discriminación y estimación de parámetros; O'Brien y Rawlings (1996) presentan una nueva propuesta para discriminar entre varios modelos competitivos no anidados y estimar simultáneamente los parámetros del modelo a partir de una combinación convexa de los logaritmos de los criterios considerados; recientemente Atkinson (2008) presenta el criterio  $\mathbb{DT}$ -optimalidad que en esencia es una combinación convexa de los criterios para discriminar,  $\mathbb{T}$ -optimalidad, y para estimar,  $\mathbb{D}$ -optimalidad. En el caso más general, también se encuentran varios trabajos para la construcción de diseños óptimos de objetivos múltiples, por ejemplo, Huang y Wong (1998), Clyde y Chaloner (1996).

En este trabajo se tiene como objetivo la construcción de un criterio de optimalidad general que tenga en cuenta múltiples objetivos tales como la discriminación entre dos modelos no lineales, la estimación de los parámetros en cada uno de los modelos y la estimación de funciones no lineales de los parámetros,  $H_j(\theta)$ . Los trabajos antes mencionados no consideran la estimación de funciones no lineales, sólo la estimación de los parámetros. Este tipo de problemas surgen de forma natural en el área de farmacocinética, allí es de interés, además de estimar el vector de parámetros asociado a un modelo de compartimientos, estimar varias funciones no lineales de éste. Estas funciones son de utilidad para entender la cinética de un medicamento bajo estudio;

entre otras están: el área bajo la curva de concentración ( $ABC$ ), el tiempo donde se alcanza la concentración máxima ( $t_{m\acute{a}x}$ ) y el valor de la concentración máxima ( $\eta(t_{max})$ ). También estas funciones son usadas para realizar estudios de bioequivalencia entre un medicamento genérico y un nuevo medicamento, Berger y Hsu (1996).

En esta investigación se construirá un criterio de optimalidad que da solución al problema planteado. El criterio de optimalidad propuesto representa un balance entre el proceso de discriminación y el de estimación tanto de los parámetros como de las funciones no lineales de interés. Con la teoría de los diseños óptimos se enunciará y demostrará el respectivo teorema de equivalencia. Además, se ilustrará el criterio de optimalidad con dos modelos de compartimientos de uso frecuente en el área de farmacocinética. Se hará un primer análisis bajo el supuesto de independencia de las observaciones. Se construirán diseños óptimos para cada modelo individualmente. Luego, a partir de cuatro variantes del criterio, se hallarán los diseños óptimos locales para discriminar y estimar las funciones no lineales en los dos modelos considerados; también se dará una metodología para construir diseños a partir de varios diseños propuestos que satisfacen los mismos objetivos de discriminación y estimación de parámetros y funciones no lineales. Los diseños óptimos se hallan numéricamente a partir de algoritmos propuestos en la literatura para la obtención de estos (Fedorov y Hackl (1997)); adicionalmente se propone un método alternativo para hallar los diseños óptimos el cual ha mostrado, al menos en los diseños hallados en esta tesis, ser más rápido que los hallados usando los algoritmos tradicionales. Con los ejemplos, para diseños aproximados se muestra que el procedimiento alternativo funciona bien, ya que con todos los diseños obtenidos se verifica el teorema de equivalencia respectivo.

Con el fin de analizar el efecto en la especificación del valor local se encontrarán diseños óptimos promediados por varias distribuciones a priori definidas alrededor del valor local. Se harán varias comparaciones con respecto a la  $\phi$ -eficiencia de los diseños obtenidos.

También, se considera la solución del mismo problema en el escenario donde se tiene una estructura de covarianza particular para las observaciones, se compararán los diseños allí obtenidos con los respectivos diseños hallados bajo el supuesto de independencia.

Este trabajo está dividido en tres partes. En la primera parte, capítulo 2, se dan algunos elementos de uso en la teoría de farmacocinética, además de un desarrollo teórico de los modelos de compartimientos. Ambos temas están estrechamente ligados y son de relevante importancia en el último capítulo de esta tesis, cuando se ilustra la propuesta en dos modelos de compartimientos.

En la segunda parte, capítulo 3, se presenta un resumen de la teoría de los diseños óptimos tanto para el caso lineal como no lineal. Se construyen diseños óptimos para varios ejemplos presentados en la literatura y se complementan con otros diseños propuestos; además se da un enfoque unificado de los trabajos de Whittle (1973), Kiefer (1974) y Silvey y Titterton (1974) relativos a los teoremas de equivalencia. Allí se presenta en forma clara y detallada tanto el enunciado como la demostración de los teoremas de equivalencia para los criterios usados en esta tesis.

La última parte comprendida por los Capítulos 4 y 5 hacen parte de la principal contribución de este trabajo. En el capítulo 4 se presenta la construcción del criterio de optimalidad que da solución al problema de hallar diseños óptimos para discriminación y estimación tanto del vector de parámetros como de funciones no lineales de éstos. Se enuncia y demuestra el teorema de equivalencia asociado al criterio de optimalidad propuesto. Además, se ilustra el criterio de optimalidad con dos modelos de compartimientos. Primero, se hace un análisis de los diseños óptimos obtenidos para cada uno de los modelos individualmente. Luego, a partir de cuatro variantes del criterio se hallan diseños tanto óptimos locales como promediados por una distribución a priori. También se llevan a cabo varias comparaciones con respecto a la  $\phi$ -eficiencia de los diseños obtenidos.

Además, en el capítulo 5 se contrastan los diseños óptimos con aquellos que se obtienen cuando se supone una estructura de correlación en los errores del tipo exponencial, igualmente se hallan diseños óptimos locales exactos y se calcula la eficiencia de los diseños aproximados con respecto al criterio de optimalidad al considerar correlación. Lo anterior se hace bajo la generalización *ad hoc* del criterio  $\mathbb{T}$ -optimalidad para observaciones correlacionadas.

En el capítulo 6 se presentan las diferentes conclusiones obtenidas de este trabajo y algunas líneas de investigación que se proponen.

Al final de la tesis se anexan cuatro trabajos que se presentaron en diferentes eventos. Los dos primeros anexos fueron documentos editados de dos posters presentados en el congreso *Moda 8, Model Oriented Data Analysis* en junio de 2007; Almagro, España y en el evento: *First Canada and Mexico Statistics Meeting* en febrero de 2008, CIMAT, Guanajuato, Gto. México. Los últimos dos anexos son publicaciones hechas en la *Revista Colombiana de Estadística*, 2007, sobre una introducción a los diseños óptimos; y en las *Memorias del XXI Foro Nacional de Estadística*, 2007, donde se hace una exposición somera de los diseños experimentales óptimos en un modelo de compartimientos.

# Capítulo 2

## Farmacocinética y modelos de compartimientos

### 2.1. Introducción

En este capítulo se presentan varios conceptos básicos del área de farmacocinética (PK, por sus siglas en Inglés; *pharmacokinetics*) que ayudan a entender el proceso de los modelos de compartimientos involucrados en el ejemplo del capítulo 4. Allí se ilustra el criterio de optimalidad propuesto en el escenario de farmacocinética. Además, para dar mayor completez a la teoría que se desarrolla se presentan algunos aspectos relacionados con los modelos de compartimientos. Estos últimos son ampliamente usados en farmacocinética, por ello se detalla la forma como se hallan las ecuaciones diferenciales involucradas en un modelo de compartimientos y adicionalmente su solución.

### 2.2. Conceptos de farmacocinética

La farmacocinética trata de entender el modo en que actúan y se comportan los medicamentos en el organismo y de predecir la forma en que actuarían en situaciones nuevas, Clark y Smith (1989). Es una disciplina que describe la evolución en el tiempo de un medicamento en el organismo y los procesos que lo afectan (las fuerzas que actúan sobre él). Hay dos fases asociadas a la cinética de un medicamento, la fase de absorción y la fase de distribución en el organismo.

La fase de absorción ocurre cuando los medicamentos son transportados por la sangre hasta su lugar de acción y cuando el medicamento es administrado por vía extravasal éste tiene que ser absorbido para poder estar presente en el torrente circulatorio. Por lo tanto, la absorción debe tener lugar antes que aparezca el medicamento en

## CAPÍTULO 2. FARMACOCINÉTICA Y MODELOS DE COMPARTIMIENTOS 6

circulación tras la administración extravasal. Lo mismo ocurre con aquellos medicamentos administrados por otras vías como: intramuscular, intraperitoneal, tópica, oral o rectal. Mientras que en los medicamentos administrados por vía intravenosa e intraarterial la absorción es instantánea.

Por otra parte, la fase de distribución ocurre cuando el medicamento se encuentra en la sangre y penetra a los órganos y tejidos del organismo con el fin de poder actuar. Es de observar que los medicamentos no suelen ser específicos de un determinado tejido de forma que alcanzan varios tejidos y órganos.

A continuación se listan algunos conceptos usados en farmacocinética:

- “Bolus”, consiste en la administración de un medicamento en forma de inyección intravenosa rápida.
- “Volumen de distribución”, denotado por  $V_D$ . Es la cantidad total de medicamento en el cuerpo dividida por la concentración en la sangre.
- “Cantidad total de medicamento” al tiempo  $t$ , está dada por:  $M_t = C_p(t)V_D$ , donde  $C_p(t)$  es la “concentración en el plasma” al tiempo  $t$ .
- “Semivida” o  $T_{\frac{1}{2}}$ , es el tiempo necesario para que la concentración de medicamento en la sangre o plasma se reduzca a la mitad.
- “Constante de velocidad de eliminación” ( $k_{el}$ ), es una constante de proporcionalidad, representa la fracción de medicamento existente a cada tiempo que es eliminada en la unidad de tiempo.
- “Aclaramiento sanguíneo”,  $Cl_s$ , y “Aclaramiento plasmático”,  $Cl_p$ . Es el volumen de sangre o plasma depurado de medicamento en la unidad de tiempo. El aclaramiento está relacionado con el volumen en el que se disuelve el medicamento (volumen de distribución) y con la velocidad a la que sale (es decir, con  $T_{1/2}$  o con  $k_{el}$ ). Se define como el producto del volumen de distribución por la constante de velocidad de eliminación,

$$Cl_p = V_D * k_{el}.$$

La velocidad a la cual se elimina el medicamento se puede hallar por:  $Cl_p * C_p$ .

- “Extracción”, se produce extracción cuando algún órgano toma, en alguna medida, medicamento de la sangre. Por extracción se indica, normalmente, la eliminación irreversible de medicamento, ya sea por excreción o por metabolismo.

## CAPÍTULO 2. FARMACOCINÉTICA Y MODELOS DE COMPARTIMIENTOS 7

Razón de extracción, es un número entre cero y uno, y se define por la siguiente relación:

$$E = \frac{C_{\text{sangre entrada}} - C_{\text{sangre salida}}}{C_{\text{sangre entrada}}}.$$

La razón de extracción se relaciona con el aclaramiento a partir de:

$$Cl = QE,$$

donde  $Q$ : es el flujo de sangre.

- “Efecto de primer paso”, se presenta cuando una gran proporción del medicamento que es absorbido va a ser eliminado antes de que llegue a circulación sistémica.
- “Área bajo la curva de concentración”:  $A(0, \infty) = \int_0^{\infty} C_p(t) dt$ .
- “Tiempo para la concentración máxima”,  $t_{\text{máx}}$ , tiempo donde la concentración alcanza su máximo valor.
- “Concentración máxima”,  $C_p(t_{\text{máx}})$ , es el máximo valor de la concentración.

Las últimas tres cantidades son estimadas en la mayoría de los estudios de farmacocinética y dan información farmacocinética del medicamento bajo estudio. También son de utilidad para realizar estudios de bioequivalencia entre dos medicamentos, un medicamento nuevo que se va a introducir al mercado con un medicamento genérico, ver Berger y Hsu (1996).

### 2.3. Modelos de Compartimientos

En esta sección se expone la teoría concerniente con modelos de compartimientos, tópico importante en el desarrollo del actual proyecto de investigación, en los ejemplos ilustrativos usados para la búsqueda de diseños óptimos con el criterio que se propone en el capítulo 4. Se empieza dando la noción de compartimiento, la forma de cómo establecer las ecuaciones diferenciales asociadas al modelo de compartimientos, y al final, para algunos casos, se presenta la solución explícita de estas ecuaciones.

Una clase de modelos no lineales de especial importancia resulta al considerar la respuesta de un fenómeno descrito por un sistema lineal de ecuaciones diferenciales ordinarias como resultado de un análisis compartimental. Un sistema compartimental se entiende como aquel que está formado por un número finito de subsistemas macroscópicos, llamados compartimientos, cada uno de los cuales es homogéneo y está bien mezclado, además hay intercambio de materiales entre los compartimientos y

## CAPÍTULO 2. FARMACOCINÉTICA Y MODELOS DE COMPARTIMIENTOS 8

el medio. Allí pueden haber entradas (inputs) del medio en uno o más compartimientos y también pueden existir salidas de uno o más compartimientos al medio, ver Seber y Wild (1989).

Un uso común de los modelos de compartimientos es en el área de farmacocinética, visto brevemente en la sección anterior. Allí se estudia el intercambio de sustancias, por ejemplo un medicamento, en un sistema biológico dividido en compartimientos. Se asume que las tasas de flujo de las sustancias entre compartimientos sigue una cinética de primer orden, esto es la tasa de transferencia de un compartimiento receptor es proporcional a la concentración en el compartimiento proveedor, o fuente. Los coeficientes de transferencia, asumidos constantes con respecto al tiempo, son llamados tasas constantes. Estos modelos son usados en diferentes ramas de la medicina y en muchas otras de la biología, incluyendo el estudio de ecosistemas en ecología. También son de utilidad en química y bioquímica, en el estudio del movimiento de poblaciones y en la dispersión de epidemias, entre otros. Por lo anterior es interesante tener técnicas especiales para hacer su análisis más tratable. A continuación se da una visión global de los modelos de compartimientos y un método para la solución del sistema de ecuaciones diferenciales.

### 2.3.1. Modelos de compartimientos y diagramas de sistemas

Los compartimientos son descritos por círculos y el intercambio de sustancia de un compartimiento a otro por una flecha, como se muestra en los dos diagramas de la Figura 2.1.

La Figura 2.1 (a) muestra un diagrama de dos compartimientos usado para modelar la cinética de un medicamento cuando se administra vía oral. El segundo compartimiento tiene una tasa de absorción de  $\gamma_{21}$ , y una tasa de eliminación al exterior de  $\gamma_{02}$ . El primer compartimiento representa el lugar de administración, el estómago en este caso. La Figura 2.1 (b) ilustra un sistema de tres compartimientos que sirve para modelar la cinética de un medicamento administrado vía oral en el tiempo  $t = 0$  y luego se distribuye en el organismo. Como en el modelo anterior, el primer compartimiento representa el sitio de administración. El compartimiento 3, conocido como compartimiento central, representa el plasma (sangre) o el sistema circulatorio. En el compartimiento 2 están representados los tejidos.

En Metzler (1971) se presenta una discusión con respecto al número de compartimientos. En farmacocinética no es considerado como un compartimiento el lugar donde se administra el medicamento, así los dos modelos considerados serían monocompartimentales y bicompartimentales. Sin embargo, en adelante se tendrá en cuenta el sitio de administración como un compartimiento.

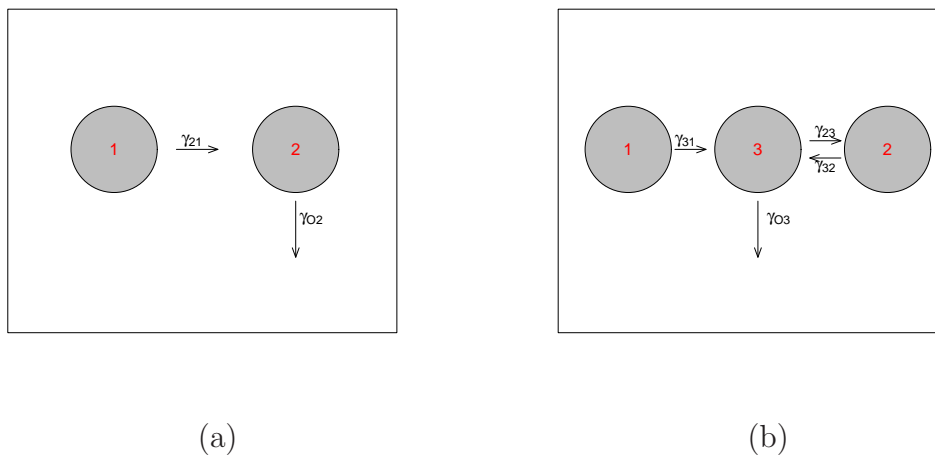


Figura 2.1: Dos modelos de compartimientos: (a) dos compartimientos, (b) tres compartimientos.

En la Figura 2.1(b) se tiene un flujo de sustancias del compartimiento 1 al 2, un intercambio entre los compartimientos 2 y 3, y una salida al medio del compartimiento central, (3).

El sistema se dice que es cerrado si no tiene intercambio con el medio, en caso contrario el sistema es abierto. Los compartimientos no necesitan ser físicamente separados. Por ejemplo en el caso de reacciones químicas, los compartimientos pueden hacer referencia a cantidades de diferentes sustancias dentro del mismo recipiente, y el “intercambio de sustancia” a la forma como ciertos componentes reaccionan para formar otros y viceversa. Cabe señalar en este punto que existen varios trabajos de diseños óptimos en esta dirección, ver por ejemplo: Ucinski y Bogacka (2005).

Los modelos de compartimientos lineales son de interés para este trabajo, aquellos donde las tasas de transferencia de sustancias del compartimiento  $j$  al compartimiento  $k$  son proporcionales a la cantidad de sustancia que está presente en el compartimiento  $j$ . Este supuesto es conocido como cinética de primer orden. Estos modelos conducen a sistemas lineales de ecuaciones diferenciales.

Siguiendo el desarrollo histórico de este tema, inicialmente se comienza con un tratamiento determinístico y después se introducen las consideraciones estocásticas en el modelo.



**Modelos lineales determinísticos**

Sea  $\eta_j(t)$  la cantidad de sustancia presente en el compartimiento  $j$  al tiempo  $t$ . Sea  $\gamma_{j0}$  la tasa de entrada de la sustancia del medio al compartimiento  $j$ . El coeficiente de transferencia fraccional,  $\gamma_{jk}$ , es la tasa de distribución desde el compartimiento  $k$  al compartimiento  $j$ , dividido por la cantidad de sustancia en  $k$ ,  $j = 0$  corresponde al medio.

Heurísticamente, en un intervalo de tiempo infinitesimal  $[t, t + dt)$  pasa una cantidad de sustancia dada por  $\gamma_{jk}\eta_k dt$  desde el compartimiento  $k$  al compartimiento  $j$ . Análogamente, una cantidad  $\gamma_{0j}dt$  pasa desde el compartimiento  $j$  al medio. Así en la Figura 2.1(b), el cambio en la cantidad de sustancia en el compartimiento 3 es:

$$d\eta_3(t) = \gamma_{31}\eta_1(t)dt - \gamma_{03}\eta_3(t)dt - \gamma_{23}\eta_3(t)dt + \gamma_{32}\eta_2(t)dt$$

y la tasa de cambio está dada por:

$$\frac{d\eta_3(t)}{dt} = \gamma_{31}\eta_1(t) - \gamma_{03}\eta_3(t) - \gamma_{23}\eta_3(t) + \gamma_{32}\eta_2(t).$$

Así, el sistema completo de la Figura 2.1(b), se describe por el siguiente conjunto de ecuaciones diferenciales:

$$\begin{aligned} \dot{\eta}_1(t) &= -\gamma_{21}\eta_1(t) \\ \dot{\eta}_2(t) &= \gamma_{23}\eta_3(t) - \gamma_{32}\eta_2(t) \\ \dot{\eta}_3(t) &= \gamma_{31}\eta_1(t) - (\gamma_{03} + \gamma_{23})\eta_3(t) + \gamma_{32}\eta_2(t). \end{aligned} \tag{2.1}$$

En forma análoga el sistema de la Figura 2.1(a), se describe por:

$$\begin{aligned} \dot{\eta}_1(t) &= -\gamma_{21}\eta_1(t) \\ \dot{\eta}_2(t) &= \gamma_{21}\eta_1(t) - \gamma_{02}\eta_2(t). \end{aligned} \tag{2.2}$$

En general, las tasas fraccionales  $\gamma_{jk}$  pueden depender del tiempo  $t$ , el estado  $\eta$  del sistema, y algunos parámetros desconocidos  $\theta$ . Además, pueden existir  $c$  compartimientos, con sustancia siendo transferida entre todos los compartimientos y entre el medio y cualquier compartimiento. Un modelo así de general se puede describir por el siguiente sistema de ecuaciones diferenciales:

$$\dot{\eta}_j(t) = \sum_{k=1}^c \gamma_{jk}(\eta, t, \theta) \eta_k(t) - \sum_{k=0}^c \gamma_{kj}(\eta, t, \theta) \eta_j(t) + \gamma_{j0}(t), \quad (j = 1, 2, \dots, c) \tag{2.3}$$

conocidas como ecuaciones de balance. En esta sección sólo se abordará el caso de sistemas lineales, donde  $\gamma_{jk}(\eta; t) \equiv \gamma_{jk}$  y  $\gamma_{j0}(t) \equiv \gamma_{j0}$ , se suponen constantes. Así, el

sistema de ecuaciones 2.3 se puede escribir en forma compacta como:

$$\dot{\eta}_j = \sum_{k=1}^c \alpha_{jk} \eta_k + \gamma_{j0},$$

donde:  $\alpha_{jk} = \gamma_{jk}; j \neq k$  y

$$\alpha_{jj} = - \sum_{\substack{k=0 \\ k \neq j}}^c \gamma_{kj} = - \sum_{\substack{k=0 \\ k \neq j}}^c \alpha_{kj}.$$

La forma matricial del sistema anterior es la siguiente:

$$\dot{\eta}(t) = A\eta(t) + r(t), \quad (2.4)$$

donde  $A = [(\alpha_{jk})]$  es una matriz  $c \times c$ . En general los  $\gamma_{jk}$ 's distintos de cero forman parte del vector  $\theta$  de parámetros desconocidos. El estado inicial del sistema  $\eta(0)$  puede ser conocido o tener parámetros desconocidos.

Los datos experimentales que son usados para ajustar el modelo de compartimientos típicamente consiste de medidas en  $\eta_j(t)$  tomadas para uno o más compartimientos en una sucesión de puntos de tiempos discretos. En la mayoría de las situaciones experimentales es imposible tomar medidas en todos los compartimientos. En algunos casos las únicas medidas disponibles son  $\sum_j \eta_j(t)$ , la cantidad total de sustancia en el sistema, o sólo se tienen las medidas en un sólo compartimiento, en el central, por ejemplo. El último caso es objeto de estudio en el capítulo 4. Pero se está consciente de que la inhabilidad o incapacidad de tomar medidas en todos los compartimientos puede causar problemas de identificabilidad con el vector de parámetros  $\theta$ , Seber y Wild (1989) presenta una discusión al respecto.

El propósito de un análisis compartimental es usualmente la estimación de  $\theta$  y/o probar el ajuste de un modelo postulado con datos experimentales o elegir entre modelos competitivos.

En la próxima sección se presenta la solución al sistema de ecuaciones diferenciales y las respectivas soluciones para los dos modelos considerados.

### 2.3.2. Solución del sistema de ecuaciones diferenciales

La solución general de (2.4) se escribe como:

$$\eta(t) = e^{At}\eta_0 + e^{At} * r(t) \quad (2.5)$$

CAPÍTULO 2. FARMACOCINÉTICA Y MODELOS DE COMPARTIMIENTOS 12

donde la matriz exponencial  $e^{At}$  representa la serie de potencias convergente:

$$e^{At} = I + \frac{At}{1!} + \frac{(At)^2}{2!} + \dots$$

donde  $*$  denota la función convolución,

$$e^{At} * r(t) = \int_0^t e^{A(t-u)} r(u) du.$$

La función vectorial se integra componente a componente. Si  $A$  es diagonalizable, entonces existe una matriz no singular de vectores propios,  $U$ , y una matriz diagonal de valores propios,  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_c)$  tal que

$$A = U\Lambda U^{-1}$$

entonces,

$$e^{At} = Ue^{\Lambda t}U^{-1},$$

siendo

$$e^{\Lambda t} = \text{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_c t}).$$

Se puede verificar que la ecuación 2.5 es solución al sistema cuando  $r(t) = 0$ ,  $\eta(0) = \eta_0$  y la matriz  $A$  es diagonalizable. En efecto, a partir de la siguiente igualdad:

$$\dot{\eta} = A\eta = U\Lambda U^{-1}\eta,$$

premultiplicando ambos lados por  $U^{-1}$ , y haciendo  $\zeta = U^{-1}\eta$ , se obtiene:

$$\dot{\zeta} = \Lambda\zeta,$$

dando un conjunto de ecuaciones diferenciales de primer orden independientes:

$$\dot{\zeta}_k = \lambda_k \zeta_k, \quad k = 1, 2, \dots, c$$

con soluciones:

$$\zeta_k(t) = e^{\lambda_k t} \zeta_k(0),$$

donde  $\zeta_k(0)$  es el  $k$ -ésimo elemento de  $U^{-1}\eta_0$ . Reinvirtiendo a  $\eta = U\zeta$ , se tiene

$$\eta(t) = Ue^{\Lambda t}U^{-1}\eta_0 = e^{At}\eta_0. \quad (2.6)$$

Si la matriz  $A$  no es diagonalizable, y/o bajo consideraciones más generales, se debe resolver el sistema de ecuaciones diferenciales por métodos numéricos (Jennrich y Brighth (1976)). Igualmente se puede usar integración numérica para obtener los gradientes, con respecto al vector de parámetros  $\theta$ , de la solución obtenida para un compartimiento particular. Este no es el caso con los modelos de compartimientos que se usarán en el capítulo 4.

CAPÍTULO 2. FARMACOCINÉTICA Y MODELOS DE COMPARTIMIENTOS 13

**Ejemplo 2.3.1.** En esta sección se van a hallar las soluciones de los modelos de compartimientos de la Figura 2.1.

El sistema de ecuaciones diferenciales de la ecuación 2.2 se representa en forma matricial como:

$$\dot{\eta}(t) = A_1\eta(t) + r_1(t),$$

donde  $\eta^T = (\eta_1, \eta_2)$ ,  $A_1 = \begin{bmatrix} -\gamma_{21} & 0 \\ \gamma_{21} & -\gamma_{02} \end{bmatrix}$  y  $r_1^T(t) = (0, 0)$ .

No es difícil mostrar que los valores propios de  $A_1$  son  $\lambda_1 = -\gamma_{21}$ ,  $\lambda_2 = -\gamma_{02}$ , con vectores propios asociados:

$$v_1^T = (\gamma_{02} - \gamma_{21}, \gamma_{21}), \quad v_2^T = (0, 1).$$

Entonces, haciendo  $U = [v_1 \ v_2]$  y  $e^{\Lambda t} = \text{diag}(e^{-\gamma_{21}t}, e^{-\gamma_{02}t})$ , se obtiene como solución del sistema anterior:

$$\eta(t) = e^{A_1 t} \eta_0 = \frac{1}{\gamma_{02} - \gamma_{21}} \begin{bmatrix} (\gamma_{02} - \gamma_{21}) e^{-\gamma_{21}t} & 0 \\ \gamma_{21} (e^{-\gamma_{21}t} - e^{-\gamma_{02}t}) & (\gamma_{02} - \gamma_{21}) e^{-\gamma_{02}t} \end{bmatrix} \eta_0,$$

y si  $\eta_0^T = (C, 0)$ , se obtiene:

$$\eta(t) = \begin{bmatrix} C e^{-\gamma_{21}t} \\ \frac{\gamma_{21}}{\gamma_{02} - \gamma_{21}} (e^{-\gamma_{21}t} - e^{-\gamma_{02}t}) \end{bmatrix}.$$

La solución obtenida para el segundo compartimiento corresponde a uno de los ejemplos dados en la sección 3.6.2 de modelos no lineales siendo  $\theta_1 = \gamma_{21}$  y  $\theta_2 = \gamma_{02}$ .

Análogamente se halla la solución para el segundo modelo de tres compartimientos. El sistema de ecuaciones dado en la ecuación 2.1 se representa por:

$$\dot{\eta}(t) = A\eta(t) + r(t),$$

siendo  $\eta(t)^T = (\eta_2(t), \eta_3(t))$ ,  $A = \begin{bmatrix} -\gamma_{32} & \gamma_{23} \\ \gamma_{32} & -(\gamma_{03} + \gamma_{23}) \end{bmatrix}$ , y  $r(t)^T = (0, A_0\gamma_{31} \exp(-\gamma_{31}t))$ .

Lo anterior se obtiene al resolver la primera ecuación diferencial, con  $\eta_1(0) = A_0$ . Acá la matriz  $A$  es cuadrada y diagonalizable, donde:

$$A = U\Lambda U^{-1}, \quad U = \begin{bmatrix} \frac{1}{\lambda_1 + \gamma_{32}} & \frac{1}{\gamma_{32}} \\ \frac{1}{\gamma_{23}} & \frac{1}{\lambda_2 + \gamma_{03} + \gamma_{23}} \end{bmatrix}, \quad \Lambda = \text{diag}(\lambda_1, \lambda_2),$$

con  $\lambda_1$  y  $\lambda_2$  soluciones de la ecuación cuadrática:  $\lambda^2 + (\gamma_{32} + \gamma_{23} + \gamma_{03})\lambda + \gamma_{32}\gamma_{23} = 0$ . Luego, se obtiene la siguiente expresión para  $e^{At}$ :

CAPÍTULO 2. FARMACOCINÉTICA Y MODELOS DE COMPARTIMIENTOS 14

$$e^{At} = \begin{bmatrix} \frac{e^{\lambda_1 t}}{\lambda_1 + \gamma_{32}} & \frac{e^{\lambda_2 t}}{\gamma_{32}} \\ \frac{e^{\lambda_1 t}}{\gamma_{23}} & \frac{e^{\lambda_2 t}}{\lambda_2 + \gamma_{03} + \gamma_{23}} \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_2 + \gamma_{03} + \gamma_{23}} \frac{1}{\Delta} & -\frac{1}{\gamma_{32}} \frac{1}{\Delta} \\ -\frac{1}{\gamma_{23}} \frac{1}{\Delta} & \frac{1}{\lambda_1 + \gamma_{32}} \frac{1}{\Delta} \end{bmatrix},$$

con

$$\frac{1}{\Delta} = \frac{\gamma_{32}\gamma_{23}(\lambda_1 + \gamma_{32})(\lambda_2 + \gamma_{03} + \gamma_{23})}{\gamma_{32}\gamma_{23} - (\lambda_1 + \gamma_{32})(\lambda_2 + \gamma_{03} + \gamma_{23})}.$$

Y con las condiciones iniciales:  $\eta_1(0) = A_0, \eta_2(0) = \eta_3(0) = 0$ , la solución está dada por:

$$\eta(t) = \int_0^t e^{A(t-s)} r(s) ds,$$

de donde al integrar término a término la matriz anterior se obtiene la siguiente solución:

$$\eta_2(t) = \frac{\gamma_{31}}{\gamma_{32}} \frac{A_0}{\lambda_1 + \gamma_{32}} \frac{1}{\Delta} \left\{ \left[ \frac{1}{\gamma_{31} + \lambda_1} - \frac{1}{\gamma_{31} + \lambda_2} \right] e^{-\gamma_{31}t} - \frac{1}{\gamma_{31} + \lambda_1} e^{\lambda_1 t} + \frac{1}{\gamma_{31} + \lambda_2} e^{\lambda_2 t} \right\}$$

$$\eta_3(t) = \frac{\gamma_{31} A_0}{\Delta} \left\{ \left[ \frac{1}{\gamma_{23}\gamma_{32}(\gamma_{31} + \lambda_1)} - \frac{1}{(\lambda_2 + \gamma_{03} + \gamma_{23})(\lambda_1 + \gamma_{32})(\gamma_{31} + \lambda_2)} \right] e^{-\gamma_{31}t} - \frac{e^{\lambda_1 t}}{\gamma_{23}\gamma_{32}(\gamma_{31} + \lambda_1)} + \frac{e^{\lambda_2 t}}{(\lambda_2 + \gamma_{03} + \gamma_{23})(\lambda_1 + \gamma_{32})(\gamma_{31} + \lambda_2)} \right\}$$

Este modelo de tres compartimientos es usado en el problema bajo estudio del capítulo 4.

# Capítulo 3

## Teoría de los diseños óptimos

### 3.1. Introducción

En este capítulo se exploran los conceptos de optimalidad tanto para modelos lineales como no lineales. Se introducen los criterios de optimalidad a partir de funciones y matrices de información, (Pukelsheim (1993)). Además, se construyen diseños óptimos para varios modelos y se presenta una versión completa y unificada de los teoremas de equivalencia presentados en los trabajos de Whittle (1973), Kiefer (1974) y Silvey y Titterington (1974). Se presentan las respectivas demostraciones. Parte del trabajo presentado en este capítulo fue publicado en López y Ramos (2007a).

En diversas áreas de investigación se tiene como objetivo modelar las relaciones una variable respuesta,  $Y$ , a través de  $k$ -variables explicativas,  $x^T = (x_1, x_2, \dots, x_k)$ , mediante un modelo estadístico de la forma:

$$Y = Y(x) = \eta(x; \theta) + \epsilon, \quad (3.1)$$

siendo  $\eta(x; \theta)$  una función lineal o no lineal en el vector de parámetros desconocido  $\theta \in \mathbb{R}^m$ ; y el término de error se supone que tiene media cero y varianza constante  $\sigma^2$ . Una vez que se especifica el modelo, el siguiente paso es 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 alcanza al construir un diseño en donde la elección de los niveles de los  $x_j$ 's y la frecuencia de medición de la respuesta está regido por algún criterio de optimalidad (con significado estadístico).

En la literatura aparecen varios ejemplos prácticos donde se han usado los diseños óptimos, ver Atkinson (1996). Además, existe un gran número de contribuciones sobre este tópico, por ejemplo, entre otros autores, Smith (1918) encontró diseños para los modelos polinomiales, Kiefer (1959) introdujo explícitamente la noción de diseño óptimo y sus propiedades; y posteriormente realizó muchos trabajos en el área, ver Brown

et al. (1985). El área de los diseños óptimos es un campo de investigación maduro donde, libros como el de Atkinson y Donev (1992) y Pukelsheim (1993), recogen los tratamientos estadístico y formal, respectivamente, de los diseños óptimos.

### 3.2. Diseños óptimos para modelos lineales

Para los modelos lineales se considera que la relación entre las  $n$ -observaciones  $Y_i$  y  $x_i$  está dada por:

$$Y_i = Y(x_i) = \theta^T f(x_i) + \epsilon_i, \quad x_i \in \mathbb{R}^k, \quad \theta \in \mathbb{R}^m, \quad (3.2)$$

donde  $f = [f_1, \dots, f_m]^T$  es un vector de  $m$ -funciones de regresión continuas, de valor real, linealmente independientes definidas en un conjunto  $\chi$ . El conjunto  $\chi \subseteq \mathbb{R}^k$  contiene todos los posibles puntos donde las observaciones pueden ser tomadas y es llamado el espacio de diseño o rango de regresión. Se supone que  $\chi$  es un subconjunto compacto de algún espacio euclidiano, con  $\theta \in \mathbb{R}^m$  un vector de  $m$ -parámetros desconocidos y  $\epsilon_i$  una variable aleatoria con media  $\mathbb{E}[\epsilon_i] = 0$  y varianza  $\text{Var}(\epsilon_i) = \sigma^2 > 0$ . Además, en la primera parte de este trabajo se supone no correlación en las  $n$ -observaciones, es mas los  $\epsilon_i$  son variables aleatorias independientes e idénticamente distribuídas.

Para la estimación puntual, el supuesto de los dos primeros momentos de la variable aleatoria son suficientes;

$$\mathbb{E}[Y|x] = \theta^T f(x), \quad \text{Var}[Y|x] = \sigma^2 > 0.$$

Sin embargo, para el problema de la construcción de regiones de confianza y pruebas de hipótesis, se supone que la respuesta  $Y$  en el punto  $x$  se distribuye normal con media  $\theta^T f(x)$  y varianza  $\sigma^2 > 0$ , es decir:

$$Y \sim N(\theta^T f(x), \sigma^2).$$

Los  $x_i$  no son necesariamente distintos, se pueden hacer observaciones repetidas en algún  $x_i$ ; además, en lo que sigue se supone que todas las observaciones no están correlacionadas:

$$\mathbb{E}[\epsilon_i \epsilon_j] = \begin{cases} \sigma^2 & i = j, \\ 0 & i \neq j. \end{cases}$$

Si las diferentes respuestas y los errores se disponen en vectores  $Y = (Y_1, \dots, Y_n)^T$  y  $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ , entonces el modelo 3.2 se puede escribir en forma matricial como:

$$Y = F\theta + \epsilon,$$

donde  $F = [f(x_1), \dots, f(x_n)]^T \in \mathbb{R}^{n \times m}$ , denota la matriz de diseño de orden  $n \times m$ . La esperanza y la matriz de varianzas covarianzas del vector  $Y$ , conocida también con el nombre de matriz de dispersión, se puede escribir como:

$$\mathbb{E}[Y] = F\theta, \quad D[Y] = \sigma^2 I_n, \quad (3.3)$$

donde  $I_n$  es la matriz identidad de orden  $n \times n$ . Siguiendo a Pukelsheim (1993), la ecuación 3.3 se conoce como el modelo lineal con los supuestos de momentos. Si adicionalmente, el supuesto de normalidad se cumple para  $Y_1, \dots, Y_n$ , entonces  $Y \sim N_n(F\theta, \sigma^2 I_n)$ , y en este caso se denomina el modelo lineal con el supuesto de normalidad.

Un hecho importante del análisis estadístico de la relación funcional 3.3 es la estimación de los parámetros desconocidos  $\theta$  y  $\sigma^2$  a partir de los datos observados  $Y = (Y_1, \dots, Y_n)^T$ . La inferencia en los modelos lineales, por el momento, se va a restringir a estimadores lineales insesgados para  $\theta$ , esto es,

$$\hat{\theta}_L = LY,$$

donde  $L \in \mathbb{R}^{m \times n}$  es una matriz dada y

$$\mathbb{E}[\hat{\theta}_L] = LF\theta = \theta, \quad \forall \theta \in \mathbb{R}^m. \quad (3.4)$$

Con el propósito de comparar diferentes estimadores lineales insesgados, se introduce el concepto de orden de Loewner.

**Definición 3.2.1.** Sea  $NND(m)$  ( $PD(m)$ ) el conjunto de las matrices definidas no negativas (positivas) de orden  $m \times m$  y  $Sim(m)$  el conjunto de las matrices simétricas de orden  $m \times m$ . Se define para  $A, B \in Sim(m)$ ,

$$A \geq B \quad \text{si y sólo si} \quad (A - B) \in NND(m), \quad A > B \quad \text{si y sólo si} \quad (A - B) \in PD(m).$$

El orden definido en el conjunto  $Sim(m)$  es un orden parcial, es decir, es antisimétrico:  $A \geq B$  y  $B \geq A \Rightarrow A = B$ ; reflexivo ( $A \geq A$ ) y transitivo:  $A \geq B$  y  $B \geq C \Rightarrow A \geq C$ . Este orden se denomina el orden de Loewner. El orden anterior no es total, pues no permite que cualesquiera dos matrices  $A$  y  $B$  en  $Sim(m)$  sean comparables.

La matriz de dispersión de un estimador lineal es definida no negativa, es decir,

$$D[\hat{\theta}_L] = D[LY] = \sigma^2 LL^T \geq 0.$$

Esta matriz de dispersión se puede minimizar (en el orden de Loewner) con respecto a todos los estimadores lineales insesgados  $\hat{\theta}_L$  de  $\theta$ , es decir, bajo el supuesto de



$r(F) := \text{rango}(F) = m$ , entonces  $\hat{\theta}^{BLUE} = (F^T F)^{-1} F^T Y$  es el mejor estimador lineal insesgado con respecto al orden de Loewner, es decir,

$$\sigma^2 (F^T F)^{-1} = D \left[ \hat{\theta}^{BLUE} \right] \leq D \left[ \hat{\theta}_L \right],$$

para cualquier estimador lineal insesgado  $\hat{\theta}_L$  de  $\theta$ .

Hay situaciones experimentales donde uno de los objetivos del investigador se traduce en realizar contrastes o comparaciones con los parámetros del modelo. En este caso es de interés la estimación de  $q$  - combinaciones lineales de  $\theta$ , representada matricialmente como  $K^T \theta$ , donde  $K \in \mathbb{R}^{m \times q}$  es conocida y de rango completo por filas, es decir,  $r(K) = q$ . Se supone que  $K^T \theta$  es estimable, es decir, existe un estimador lineal insesgado para  $K^T \theta$ , o en forma equivalente si se cumple:

$$\mathcal{C}(K) \subseteq \mathcal{C}(F^T) = \mathcal{C}(F^T F), \quad (3.5)$$

donde  $\mathcal{C}(A)$  denota el espacio generado por las columnas de la matriz  $A$ . Además, se puede mostrar que el mejor estimador lineal insesgado (con respecto al orden de Loewner) para el subsistema parametral,  $K^T \theta$ , está dado por:

$$\hat{\theta}^K = K^T (F^T F)^- F^T Y,$$

donde  $(F^T F)^-$  denota una inversa generalizada de la matriz  $F^T F$ . La matriz de dispersión (mínima) de este estimador está dada por:  $D \left[ \hat{\theta}^K \right] = \sigma^2 K^T (F^T F)^- K$ . Se puede mostrar que bajo la condición de inclusión de rango dada en la ecuación 3.5, el estimador  $\hat{\theta}^K$  y su matriz de dispersión no dependen de la elección de la inversa generalizada de  $(F^T F)^-$ .

A continuación se va a dar la noción de diseño experimental tanto en el caso de diseños exactos como aproximados, dando mayor énfasis en este trabajo a los últimos.

Sean  $x_1, x_2, \dots, x_s$ ,  $s$ -puntos distintos, y  $n_i$  el número de repeticiones del punto  $x_i$  en la muestra. Con lo anterior se define una medida de probabilidad  $\xi_{(n)}$  en el espacio de diseño  $\chi$  con soporte finito  $\{x_1, \dots, x_s\}$  y probabilidad  $n_i/n$  para el punto  $x_i$ . Para una muestra de tamaño  $n$ , se considera un diseño exacto a cualquier medida de probabilidad con soporte finito con probabilidades proporcionales a  $1/n$ . Se denota en forma resumida esta medida por medio de la matriz:

$$\xi_{(n)} = \begin{bmatrix} x_1 & \dots & x_s \\ \frac{n_1}{n} & \dots & \frac{n_s}{n} \end{bmatrix},$$

la primera fila contiene los puntos del espacio de diseño  $\chi$  donde las observaciones han de ser tomadas y la segunda fila indica la proporción de observaciones a tomar

en cada punto.

Note que la matriz  $F^T F$  se puede escribir en función del diseño  $\xi_{(n)}$ , ya que

$$F^T F = \sum_{i=1}^s n_i f(x_i) f^T(x_i) = n \sum_{i=1}^s \frac{n_i}{n} f(x_i) f^T(x_i) = n \int_{\mathcal{X}} f(x) f^T(x) d\xi_{(n)}(x) := nM(\xi_{(n)}),$$

donde la matriz de momentos,  $M(\xi_{(n)})$ , es  $\frac{F^T F}{n}$ .

En contraste con los diseños exactos, se definen los diseños aproximados o continuos,

$$\xi = \begin{bmatrix} x_1 & \dots & x_s \\ w_1 & \dots & w_s \end{bmatrix},$$

con  $w_i = \xi(x_i)$ , siendo  $\xi$  una medida de probabilidad definida en  $\mathcal{B}$ , conjunto de Borel de  $\mathcal{X}$  que incluye los conjuntos unitarios; tal que  $\xi$  tiene soporte finito. El soporte de  $\xi$  es  $\text{Supp}(\xi) = \{x_j \in \mathcal{X} | \xi(x_j) > 0\} = \{x_1, \dots, x_s\}$ ,  $s$ : número de puntos de soporte de  $\xi$ , y las observaciones  $Y(x)$  se hacen en  $x_1, \dots, x_s$  con frecuencias (o pesos) aproximadamente proporcionales a  $w_1, \dots, w_s$ . El conjunto de todos los diseños aproximados se denota por  $\Xi$ . En lo que sigue se hace referencia a los diseños aproximados. Como se verá más adelante, estos diseños poseen la ventaja de que la búsqueda de un diseño óptimo conlleva a un problema de optimización convexa, mientras que en el caso de diseños exactos es un problema de optimización entera, el cual es en general más difícil de enfrentar. Sin embargo, como en las aplicaciones se usan diseños exactos, es necesario utilizar un método de redondeo eficiente que convierte diseños aproximados en diseños exactos, ver Pukelsheim (1993), capítulo 12. Este método se usará en el capítulo 4, Sección 5.2 de este trabajo.

Para cada diseño  $\xi$ , la versión generalizada de la matriz de momentos está dada por:

$$M(\xi) \equiv \int_{\mathcal{X}} f(x) f^T(x) d\xi(x) = \sum_{i=1}^s f(x_i) f^T(x_i) w_i = F^T W F, \quad (3.6)$$

donde  $W = \text{diag}(w_1, \dots, w_s)$ , y  $F = [f(x_1), \dots, f(x_s)]^T \in \mathbb{R}^{s \times m}$ .

La forma de como cuantificar la información suministrada por la matriz de momentos depende de la definición de algunos funcionales de valor real (con un significado estadístico) de la matriz de momentos sobre  $\Xi$ , denominados criterios de optimalidad. Estos criterios de optimalidad se presentan a continuación, siguiendo el enfoque de Pukelsheim (1993), quién define la función intermedia  $C_K$ , con el fin de dar cuenta de la “información” contenida en combinaciones lineales de  $\theta$ . Luego se dá la noción de función de información,  $\phi$ . Para cada diseño, la matriz  $C_K(M(\xi))$  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.

**Definición 3.2.2.** Sea  $\xi \in \Xi$  un diseño aproximado. Para una matriz dada  $K \in \mathbb{R}^{m \times q}$  con  $r(K) = q \leq m$ , se denomina:

$$\mathbb{A}(K) = \{\xi \in \Xi \mid \mathcal{C}(K) \subseteq \mathcal{C}(M(\xi))\},$$

el conjunto de todos los diseños factibles y se define

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

tal que:  $C_K(M(\xi)) = (K^T M^{-1}(\xi) K)^{-1}$ , con  $\xi \in \mathbb{A}(K)$ .  $C_K(M(\xi))$  se llama la matriz de información de  $\xi \in \mathbb{A}(K)$  para el subsistema parametral  $K^T\theta$ .

Note que para  $\xi$  un diseño factible,  $C_K(M(\xi))$  cumple con las siguientes propiedades: homogeneidad positiva, esto es:  $C_K(\delta A) = \delta C_K(A)$ ,  $A \geq 0$ ,  $\delta > 0$ ; superaditividad:  $C_K(A + B) \geq C_K(A) + C_K(B)$ ,  $A, B \geq 0$ ; concavidad:  $C_K((1 - \alpha)A + \alpha B) \geq (1 - \alpha)C_K(A) + \alpha C_K(B)$ ,  $A, B \in NND(m)$ ,  $0 < \alpha < 1$ ; isotonía:  $A \geq B \Rightarrow C_K(A) \geq C_K(B)$ . Además de cumplir la siguiente propiedad en cuanto al rango o imagen de esta función,  $\text{Rango}(C_K) \subseteq NND(q)$ .

Todas las propiedades anteriores hacen referencia a propiedades deseables de la forma de como cuantificar la información. La matriz  $C_K(M(\xi))$  desempeña un papel importante en la definición del problema de diseño, que se da más adelante. Observe que si  $K = I_m$ , es decir en este caso interesa estimar  $\theta$ , y si  $M(\xi)$  es no singular, entonces  $C_I(M(\xi)) = M(\xi)$ , coincidiendo la matriz de información con la matriz de momentos, por esta razón en la literatura  $M$  también se llama matriz de información.

Una vez definida la matriz de información, el siguiente paso consiste en buscar la forma de cómo cuantificar la información suministrada por cada diseño, ya sea por la matriz de momentos o la matriz de información. Esto se hace a partir de la definición de un funcional definido en el conjunto de las matrices definidas no negativas;  $\phi : NND(q) \rightarrow \mathbb{R}$ . La función  $\phi$  es de información si cumple con ser: homogéneamente 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  $\alpha \in \mathbb{R}$ . En lo que sigue  $\phi$  denotará una función de información.

Con lo anterior, ya se tienen las herramientas necesarias para formular el problema de diseño que se da a continuación:

**Definición 3.2.3.** *El problema de diseño para el sistema parametral  $K^T\theta$  consiste en encontrar un diseño  $\xi^*$  que sea factible y que maximice, sobre todos los diseños  $\xi$  factibles para  $K^T\theta$ , la función de información:*

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

Por las propiedades de  $\phi$  y  $C_K$ , principalmente la semicontinuidad superior y por la compacidad de  $\chi$  dada en la definición, el máximo anterior se alcanza para algún diseño  $\xi$  (Pukelsheim (1993)). Si  $\xi^*$  es solución del problema de diseño, entonces se dirá que  $\xi^*$  es  $\phi$ -óptimo.

En el caso de que  $K = c$ ,  $c \in \mathbb{R}^{m \times 1}$  entonces el criterio asociado se denomina  $c$ -optimalidad; se puede mostrar que la única función de información es la identidad:  $\phi(\delta) = \delta$  y el problema de diseño se reduce a encontrar un diseño  $\xi^*$  que sea factible para  $c^T\theta$  y maximice la función de información:

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

El lado derecho de la igualdad anterior es el inverso de la varianza asociada al estimador óptimo para  $c^T\theta$ ; luego los diseños  $c$ -óptimos son aquellos que minimizan la varianza de  $c^T\hat{\theta}$ . Note que en el caso de diseños  $\xi$  aproximados, el problema de diseño dado en la Definición 3.2.3 es equivalente a un problema de optimización convexo, en donde en vez de maximizar se considera minimizar el negativo de la función respectiva. Entonces se pueden hacer uso de los resultados de la teoría de funciones convexas para resolverlo. En cambio si  $\xi$  es un diseño exacto, el problema de diseño se convierte en un problema de optimización entera el cual, en general, es más difícil de resolver.

A continuación se exhibe una clase de funciones de información introducida por Kiefer, denominado media matricial de orden  $p$ , (*matrix means*), la cual da origen a los criterios de optimalidad de mayor popularidad.

**Definición 3.2.4.** *Sea  $C \in NND(q)$ , para  $C > 0$ , se define la media matricial de orden  $p$ ,  $\phi_p$ , como la siguiente función:*

$$\phi_p(C) = \begin{cases} \lambda_{\max}(C), & p = \infty \\ \left[\frac{1}{q} \text{Tr}(C^p)\right]^{1/p}, & p \neq 0, p \neq \pm\infty \\ (\det(C))^{1/q}, & p = 0 \\ \lambda_{\min}(C), & p = -\infty \end{cases} \quad (3.7)$$

donde  $\lambda_{\max}(C)$  y  $\lambda_{\min}(C)$  denotan el máximo y el mínimo valor propio de la matriz  $C$ , respectivamente.

**Observación 3.2.1.** *Anotaciones del criterio  $\phi_p$ .*

- $\phi_p$  es una función de información para  $p \in [-\infty, 1]$ .
- Si un diseño  $\xi$  maximiza el criterio anterior, con  $C = C_K(M(\xi))$ , se dice que el diseño es  $\phi_p$ -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 en su versión generalizada. Todos dependen de la maximización del respectivo funcional evaluado en la matriz de información, en algunos casos evaluado en el diseño  $\xi$ , ellos son, respectivamente,
  - $\mathbb{D}_K$ -optimalidad, ( $p = 0$ ), criterio del determinante, equivale, en el caso de diseños exactos, a minimizar el volumen del elipsoide asociado a la estimación del sistema  $K^T\theta$ , cuando los errores son normales. En este caso el elipsoide del  $(1 - \alpha)100\%$  de confianza para el vector  $K^T\theta$  está dado por el conjunto de vectores  $\mu \in \mathbb{R}^m$  que cumplen:

$$(\hat{\theta}^K - \mu)^T (K^T M^-(\xi) K)^{-1} (\hat{\theta}^K - \mu) \leq \frac{m\hat{\sigma}^2}{n} f_{m, n-r(F), 1-\alpha},$$

con  $M(\xi) = \frac{1}{n} F^T F$ ,  $\hat{\sigma}^2 = (n - r(F))^{-1} Y^T (I_n - F(F^T F)^- F^T) Y$  y  $f_{m, n-r(F), 1-\alpha}$  denota el cuantil  $1 - \alpha$  de la distribución  $f$  con parámetros  $m$  y  $n - r(F)$ . El volumen del elipsoide es proporcional a:  $\sqrt{\det(K^T M^-(\xi) K)}$ .

Si  $K = I$ , el interés está en la estimación del vector de parámetros, y en este caso el criterio asociado se denomina:  $\mathbb{D}$ -optimalidad.

No es difícil mostrar que este criterio, cuando  $K = I$ , es invariante bajo reparametrizaciones, es decir, los diseños  $\mathbb{D}$ -óptimos obtenidos para el modelo  $Y = F\theta + \epsilon$ , son los mismos que para el modelo:

$$Y_1 = F_1\theta + \epsilon_1 = FA\theta + \epsilon_1, \quad (3.8)$$

donde  $A$  es una matriz definida positiva. En efecto, el diseño  $\xi$  es  $\mathbb{D}_A$ -óptimo para estimar  $A\theta$  si maximiza:

$$\phi_0(\xi) = \left\{ \det (AM^-(\xi)A^T)^{-1} \right\}^{1/q}$$

o equivalente si minimiza

$$-\frac{1}{q} \log(\det (AM^-(\xi)A^T)) = -\frac{1}{q} \log(\det(M^-(\xi))) - \frac{1}{q} \log(\det(A^T A)),$$

donde el segundo término de la última igualdad no depende del diseño, y por tanto el diseño  $\mathbb{D}$ -óptimo para estimar  $\theta$  en el primer modelo es equivalente al diseño  $\mathbb{D}_A$ -óptimo para estimar  $A\theta$  en el modelo 3.8.

- $\mathbb{L}$ -optimalidad,, ( $p = -1$ ), criterio promedio. Es proporcional al recíproco del promedio de las varianzas asociado a las  $q$ -combinaciones lineales de  $\theta$ . Es decir:

$$\phi_{-1}(\xi) = \left[ \frac{1}{q} \text{Tr} (K^T M^{-1}(\xi) K) \right]^{-1}.$$

En el caso donde  $K = I$ , este criterio recibe el nombre de  $\mathbb{A}$ -optimalidad, promedio de las  $m$  varianzas de cada uno de los estimadores de las componentes del vector  $\theta$ , y

- $\mathbb{E}$ -optimalidad,, ( $p = -\infty$ ), criterio del valor propio, maximización del valor propio más pequeño. En este caso:

$$\begin{aligned} \phi_{-\infty}(C_K(M(\xi))) &= \lambda_{\min} \left( (K^T M^{-1}(\xi) K)^{-1} \right) \\ &= (\lambda_{\max} (K^T M^{-1}(\xi) K))^{-1} \\ &= \left\{ \max_{z \in \mathbb{R}^q, \|z\|=1} z^T K^T M^{-1}(\xi) K z \right\}^{-1}, \end{aligned}$$

es decir, el diseño  $\mathbb{E}$ -optimal minimiza la varianza más grande posible entre todas las combinaciones lineales  $z^T K^T \theta$ , con  $z$  un vector de norma uno.

El problema de optimización planteado en la Definición 3.2.3 es muy complejo. En la práctica se hace uso de teoremas de equivalencia para verificar si un diseño dado es  $\phi$ -óptimo, además de facilitar la construcción de algoritmos computacionales para la búsqueda de diseños  $\phi$ -óptimos (Pukelsheim (1993), Atkinson y Donev (1992)). El primer teorema de equivalencia lo demostraron Kiefer y Wolfowitz (1960), allí establecieron la equivalencia entre  $\mathbb{D}$ -optimalidad y  $\mathbb{G}$ -optimalidad. A propósito, un diseño  $\xi$  es  $\mathbb{G}$ -óptimo si minimiza:

$\forall \xi \in \Xi$ ,

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

donde  $d(x; M(\xi)) = f^T(x) M^{-1}(\xi) f(x)$  se conoce como la varianza de predicción. Es decir,  $\xi$  es  $\mathbb{G}$ -óptimo si minimiza la varianza más grande posible sobre  $\chi$ .

**Teorema 3.2.1.** *Teorema de Equivalencia de Kiefer - Wolfowitz.* Sea  $\chi \subseteq \mathbb{R}^k$  con  $m$ -vectores linealmente independientes. Un diseño  $\xi$  con matriz de momentos  $M(\xi)$ , definida positiva, es  $\mathbb{D}$ -óptimo si y sólo si  $\xi$  es  $\mathbb{G}$ -óptimo si y sólo si:  $\forall x \in \chi$ ,  $f^T(x) M^{-1}(\xi) f(x) \leq m$ , si y solo si  $\bar{d}(M(\xi)) = m$ .

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

Por lo popular de los criterios  $\phi_p$  ( $p \in [-\infty, 1]$ ) y considerando el caso general donde  $M \in \text{NND}(m)$ , se enuncia el siguiente teorema de equivalencia que da condiciones

necesarias y suficientes para garantizar que un diseño dado sea  $\phi_p$ - óptimo. Para la demostración se puede consultar Pukelsheim (1993), capítulo 7, página 178. En lo que sigue, por simplicidad,  $M \equiv M(\xi)$  y  $\mathcal{M}^-$  es el conjunto de las inversas generalizadas de  $M$ .

**Teorema 3.2.2.** *Sea  $\phi_p$ ,  $p \in (-\infty, 1]$ ,  $\mathcal{M}$  un subconjunto convexo y compacto de  $NND(m)$  y  $M(\xi) \in \mathcal{M}$ , con  $\xi$  factible para  $K^T\theta$  y matriz de información  $C = C_K(M(\xi))$ . Entonces:*

- $\xi$  es  $\phi_p$ -óptimo para  $K^T\theta$  en  $\mathcal{M}$  si y solo sí:  $\exists G \in \mathcal{M}^-$  tal que:

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

*En caso de optimalidad, la igualdad se obtiene si en vez de  $A$  se coloca  $M$  u otra matriz  $\tilde{M} \in \mathcal{M}$   $\phi_p$ -óptima para  $K^T\theta$  en  $\mathcal{M}$ .*

- Si  $0 < M(\xi) \in \mathcal{M}$ , entonces  $\xi$  es  $\phi_p$ -óptimo para  $\theta$  en  $\mathcal{M}$  si y solo sí:  $\text{Tr}(AM^{p-1}(\xi)) \leq \text{Tr}(M^p(\xi))$ ,  $\forall A \in \mathcal{M}$ .

Para  $p = 0$  y  $M > 0$ , la condición 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 condición para  $A$ , y el lado izquierdo de la desigualdad es:

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

mostrando un caso particular de una de las equivalencias del Teorema 3.2.1.

Para  $p = -1$  ( $\mathbb{L}$ -optimalidad),  $M > 0$  y  $C = (K^TM^{-1}K)^{-1}$ , la condición a verificar será,

$$D(x; \xi) := f^T(x)M^{-1}KK^TM^{-1}f(x) - \text{Tr}(K^TM^{-1}K) \leq 0, \quad \forall x \in \chi. \quad (3.9)$$

La expresión del lado izquierdo de la desigualdad anterior se denomina la función de varianza o de sensibilidad asociada al criterio y se denota por  $D(x; \xi)$ . Existe la versión del teorema de equivalencia para  $\mathbb{E}$ -optimalidad ( $p = -\infty$ ), ver Pukelsheim (1993).

En la sección 3.4 se presenta una versión más general del Teorema 3.2.2, con su demostración, para cualquier función de información  $\phi$  cóncava, usando la noción de derivada direccional. En esa sección se presenta un enfoque unificado de los trabajos de Whittle (1973), Kiefer (1974) y Silvey y Titterington (1974); además se verá el teorema de equivalencia de Kiefer y Wolfowitz (Teorema 3.2.1) como un caso particular.

A continuación se muestran dos ejemplos ilustrativos, donde se encuentran diseños óptimos y se hace uso de los respectivos teoremas de equivalencia.

**Ejemplo 3.2.1. Diseños óptimos para modelos polinomiales.**

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

$$Y(x) = \mathbf{f}^T(x)\theta + \epsilon,$$

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

En el caso  $\mathbb{D}$ -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  $\mathbb{D}$ -óptimo para estimar  $\theta$  (tomando  $K = I$ ).

En efecto, bastará con mostrar que el diseño  $\xi$  verifica las condiciones del Teorema 3.2.1. Primero note que su matriz de momentos es:

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

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

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

En la Figura 3.1 se muestra que la función,  $D(x; \xi) = d(x; M(\xi)) - 3$  tiene todos sus valores por debajo de cero, y en los puntos de soporte alcanza su máximo, luego  $\xi$  es  $\mathbb{D}$ -óptimo para estimar el vector de parámetros  $\theta$ .

En la literatura existe la solución para el caso general, polinomios de grado  $d - 1$ , para los diseños  $\mathbb{D}$ -óptimos en el intervalo  $[-1, 1]$  (Pukelsheim (1993), capítulo 9, página 214; Atkinson y Donev (1992), capítulo 11, página 125). Los autores usan como argumento el Teorema 3.2.1, y muestran que los diseños  $\mathbb{D}$ -óptimos tienen igual peso  $1/d$  en los puntos de soporte,  $x$ , que son solución a la ecuación:

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

donde  $\dot{P}_d(x)$  es la derivada del polinomio de Legendre de grado  $d$ , dada por la siguiente expresión:

$$P_d(x) = \sum_{i=0}^N \left[ \frac{(-1)^i (2d - 2i)! x^{d-2i}}{2^d i! (d - i)! (d - 2i)!} \right],$$

donde

$$N = \begin{cases} \frac{d}{2} & \text{si } d \text{ es par} \\ \frac{d-1}{2} & \text{si } d \text{ es impar} \end{cases}$$



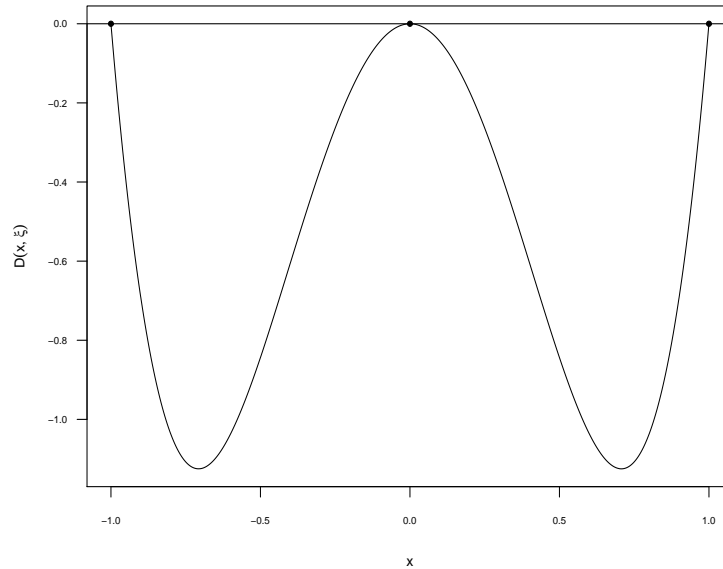


Figura 3.1: Gráfico de la función de varianza  $D(x; \xi)$ , modelo cuadrático.

Como lo observó Stigler, S. M. (1971), los diseños  $\mathbb{D}$ -óptimos en cualquier intervalo de la forma  $[a, b]$ , se pueden hallar al transformar linealmente el diseño óptimo obtenido en el intervalo  $[-1, 1]$ , esto es porque el criterio  $\mathbb{D}$ -óptimo es también invariante bajo transformaciones lineales en las variables regresoras. Hay una gran diversidad de trabajos en diseños óptimos para modelos polinomiales, por ejemplo: Hoel y Levine (1964) encontraron diseños óptimos para extrapolación; mientras que Dette, Melas y Pepelyshev (2004a) hallaron diseños óptimos para la estimación de los coeficientes individuales.

A continuación se da un ejemplo donde se tienen dos variables regresoras.

### Ejemplo 3.2.2. Modelo de regresión lineal en dos variables.

Considere el modelo de regresión lineal en dos variables:  $Y(x) = \theta_1 x_1 + \theta_2 x_2 + \epsilon$ , donde  $x_1 \in [0, a]$ ,  $x_2 \in [0, b]$ ,  $a, b > 0$ .

En este caso  $\chi = [0, a] \times [0, b]$ , y  $f^T(x) = (x_1, x_2)$ . Se va a encontrar un diseño  $\mathbb{A}$ -óptimo para estimar  $\theta^T = (\theta_1, \theta_2)$ , es decir, hallar un diseño  $\xi$  tal que minimice  $1/2 \text{Tr} M^{-1}(\xi)$ , donde

$$\xi = \begin{bmatrix} (x_{11}, x_{21})^T & \dots & (x_{1s}, x_{2s})^T \\ w_1 & \dots & w_s \end{bmatrix}.$$

Un diseño candidato tiene como puntos de soporte  $(a, 0)^T, (0, b)^T, (a, b)^T$  con pesos

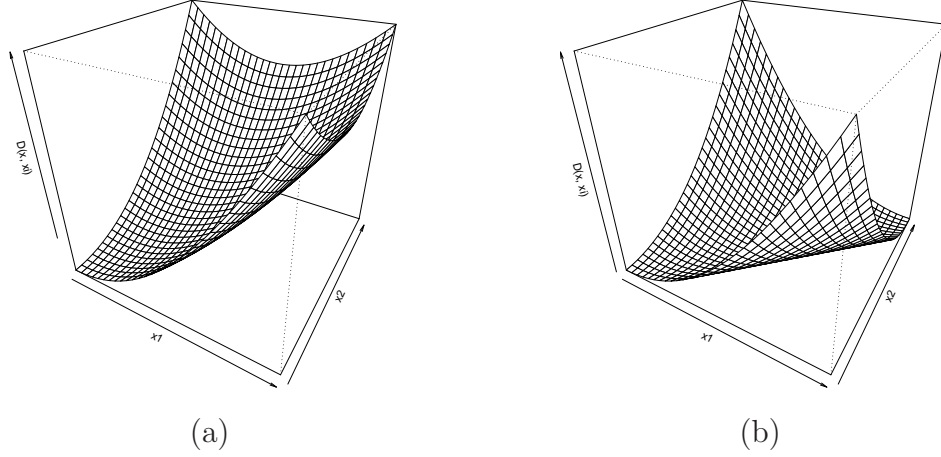


Figura 3.2: Gráfico de las funciones de varianza,  $D(x; \xi) = D\left((x_1, x_2)^T; \xi\right)$ , para los diseños: (a)  $\mathbb{A}$ -óptimo, (b)  $c$ -óptimo.

$w_1, w_2, 1 - w_1 - w_2$ . Por ejemplo, para el caso donde  $a = b = 1$ , se obtuvo que el diseño óptimo tiene pesos:  $w_1 = 0.4227, w_2 = 0.4227, w_3 = 0.1547$ , en los puntos  $(1, 0)^T, (0, 1)^T, (1, 1)^T$ . En este caso la matriz de momentos y su inversa están dadas por:

$$M(\xi) = \begin{bmatrix} a^2(1 - w_2) & ab(1 - w_1 - w_2) \\ ab(1 - w_1 - w_2) & b^2(1 - w_1) \end{bmatrix}$$

$$M^{-1}(\xi) = \frac{1}{\Delta} \begin{bmatrix} b^2(1 - w_1) & -ab(1 - w_1 - w_2) \\ -ab(1 - w_1 - w_2) & a^2(1 - w_2) \end{bmatrix},$$

donde  $\Delta = a^2b^2[w_1(1 - w_1) + w_2(1 - w_1 - w_2)]$ . Luego:

$$\text{Tr}(M^{-1}(\xi)) = \frac{b^2(1 - w_1) + a^2(1 - w_2)}{\Delta}.$$

Bastaría verificar, usando el respectivo teorema de equivalencia, que el diseño anterior es  $\mathbb{A}$ -óptimo. En efecto, haciendo  $K = I$  en la ecuación 3.9 se obtiene que la desigualdad a verificar es:

$$f^T(x)M^{-1}(\xi)M^{-1}(\xi)f(x) - \text{Tr}(M^{-1}(\xi)) \leq 0, \quad \forall x^T = (x_1, x_2) \in [0, a] \times [0, b].$$

En la Figura 3.2 (a) se muestra el caso particular para  $a = b = 1$ , y se verifica que en efecto el diseño propuesto es  $\mathbb{A}$ -óptimo.

Volviendo al caso anterior, si el interés del investigador está en comparar ambos parámetros mediante el contraste,  $\theta_1 - \theta_2$ , entonces en este caso  $c^T\theta = (1, -1)\theta$ , y

el diseño  $c$ -óptimo  $\xi$  es tal que minimiza  $c^T M^{-1}(\xi)c$ . En este caso el diseño  $c$ -óptimo obtenido para  $a = b = 1$  está dado por:

$$\xi_c = \begin{bmatrix} (1, 0)^T & (0, 1)^T \\ 0.5 & 0.5 \end{bmatrix},$$

con  $M(\xi_c) = \text{diag}(0.5, 0.5)$ .

En la ecuación 3.9 al considerar  $K = c$ , se obtiene la respectiva función de varianza o de sensibilidad dada por:

$$D(x; \xi_c) = f^T(x)M^{-1}(\xi_c)cc^T M^{-1}(\xi_c)f(x) - c^T M^{-1}(\xi_c)c.$$

En la Figura 3.2 (b) se encuentra la función anterior. Claramente  $D(x; \xi_c) \leq 0$ , para  $x^T = (x_1, x_2) \in [0, 1] \times [0, 1]$ .

### 3.3. Eficiencia de los diseños

Observe que los diseños óptimos dependen del criterio de optimalidad usado. Una forma de comparar dos diseños  $\xi_1, \xi_2$  es a partir del cálculo de las  $\phi$ -eficiencias. La  $\phi$ -eficiencia del diseño  $\xi$ , abreviada por  $\phi\text{-ef}(\xi)$ , está definida como:

$$\phi - \text{ef}(\xi) \equiv \phi - \text{eficiencia}(\xi) = \frac{\phi(\xi)}{\sup_{\xi^* \in \Xi} \phi(\xi^*)}.$$

La función  $\phi - \text{ef}(\xi)$  está entre cero y uno. Proporciona una medida de cuanto porcentaje en información está suministrando el diseño  $\xi$  con respecto a la información máxima. Así, si  $\xi_1$  fue obtenido a partir de un criterio de optimalidad  $\phi_1$ , entonces al calcular la eficiencia de cualquier otro diseño  $\xi_2$  este valor dará una idea de que tan cercano o alejado está éste de la información suministrada por el diseño  $\phi_1$ -óptimo.

Esta noción es clara en el caso de los diseños  $c$ -óptimos, pues:

$$\phi_c - \text{ef}(\xi) = \frac{(c^T M^{-1}(\xi)c)^{-1}}{\sup_{\xi^*} (c^T M^{-1}(\xi^*)c)^{-1}} = \frac{\inf_{\xi^*} c^T M^{-1}(\xi^*)c}{c^T M^{-1}(\xi)c},$$

es el cociente de dos varianzas, la varianza de  $c^T \hat{\theta}$  aportada por el diseño  $\xi$  y la varianza mínima posible.

En el caso de  $\mathbb{D}$ -optimalidad:

$$D - \text{ef}(\xi) = \phi_0 - \text{ef}(\xi) = \left[ \frac{|M(\xi)|}{\sup_{\xi^* \in \Xi} |M(\xi^*)|} \right]^{1/m},$$

es proporcional al número de puntos de soporte del diseño, Atkinson y Donev (1992). Lo anterior permite dar una interpretación de la eficiencia obtenida en función del número de observaciones requeridas para que el diseño en cuestión sea tan eficiente como el óptimo.

Antes de continuar con la teoría concerniente a los diseños óptimos para modelos no lineales, en la siguiente sección se estudian los teoremas de equivalencia, que se usarán en las secciones y capítulos siguientes.

### 3.4. Teoremas de equivalencia

En el área de los diseños óptimos son de gran relevancia los así llamados *teoremas de equivalencia*. Sirven para encontrar algorítmicamente diseños  $\phi$ -óptimos, y en algunos casos para verificar que en efecto el diseño obtenido por algún proceso de optimización numérica, es el óptimo. Se dedicará esta sección a introducir los conceptos y resultados necesarios para demostrar el teorema general de equivalencia, adicionalmente se usará el teorema para deducir los respectivos teoremas de equivalencia para  $\mathbb{D}$ ,  $\mathbb{L}$ ,  $\mathbb{D}_K$  y  $\mathbb{D}_s$  optimalidad.

En lo que sigue se denotará por:  $\chi$  el rango de regresión,  $M(\xi) = \int_{\chi} f(t) f^T(t) \xi(dt)$  la matriz de información asociada al diseño  $\xi$ ,  $\Xi$  el conjunto de todos los diseños  $\xi$  y  $\phi$  es un criterio de optimalidad cóncavo, es decir:  $\phi$  es una función de  $NND(m)$  a  $\mathbb{R}^+$ , tal que:  $\forall \alpha \in (0, 1)$ ,  $\forall A, B \in NND(m)$  se cumple:

$$\phi(\alpha A + (1 - \alpha) B) \geq \alpha \phi(A) + (1 - \alpha) \phi(B). \quad (3.10)$$

En este trabajo se considera un subconjunto de  $NND(m)$ ,  $\mathbb{M}(\Xi) = \{M(\xi) : \xi \in \Xi\}$ , el conjunto de todas las matrices de información, asociado a los diferentes diseños  $\xi \in \Xi$ .

A continuación se presentará una versión unificada de los trabajos de Whittle (1973), Kiefer (1974) y Silvey y Titterton (1974). Sean  $A$  y  $B \in NND(m)$  y  $\phi$  un criterio de optimalidad cóncavo, se define la derivada direccional en  $A$  en la dirección de  $B$  por:

$$\Phi(A; B) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \{ \phi((1 - \epsilon) A + \epsilon B) - \phi(A) \}. \quad (3.11)$$

Dados dos diseños  $\xi$  y  $\eta \in \Xi$ , se define la derivada direccional de  $\xi$  en la dirección de  $\eta$  por:

$$\Phi(\xi; \eta) \equiv \Phi(M(\xi); M(\eta)).$$

A continuación se van a enunciar y demostrar algunos resultados asociados con la función  $\Phi(\cdot; \cdot)$ .

**Lema 3.4.1.** *Sean  $A$  y  $B$  matrices definidas no negativas, entonces:*

1.  $\Phi(A; B)$  existe y  $\Phi(A; A) = 0$ .
2.  $\Phi(A; B) \geq \phi(B) - \phi(A)$ .
3. Si  $\phi$  es diferenciable, es decir, cuando:

$$\nabla\phi(A) = \frac{\partial\phi(A)}{\partial A} = \left[ \frac{\partial\phi(A)}{\partial a_{ij}} \right] = \left[ \frac{1}{2} \frac{\partial\phi(A)}{\partial a_{ij}} (1 + \delta_{ij}) \right]$$

existe, entonces:

$$\Phi(A; B) = \text{Tr}\{(B - A) \nabla\phi(A)\}. \quad (3.12)$$

**Demostración:**

Para mostrar 1, basta verificar que  $\varsigma(\epsilon) = \frac{1}{\epsilon} \{\phi((1 - \epsilon)A + \epsilon B) - \phi(A)\}$  es una función decreciente en  $(0, 1)$  y además acotada inferiormente. Veamos que  $\varsigma(\epsilon)$  es decreciente, note que, si  $0 < \epsilon_1 < \epsilon_2 < 1$ , se tiene que:

$$\begin{aligned} \phi((1 - \epsilon_1)A + \epsilon_1 B) - \phi(A) &= \phi\left(\frac{\epsilon_1}{\epsilon_2}((1 - \epsilon_2)A + \epsilon_2 B) + \left(1 - \frac{\epsilon_1}{\epsilon_2}\right)A\right) - \phi(A) \\ &\geq \frac{\epsilon_1}{\epsilon_2} \phi((1 - \epsilon_2)A + \epsilon_2 B) + \left(1 - \frac{\epsilon_1}{\epsilon_2}\right) \phi(A) - \phi(A) \\ &= \frac{\epsilon_1}{\epsilon_2} [\phi((1 - \epsilon_2)A + \epsilon_2 B) - \phi(A)] \end{aligned}$$

luego  $\varsigma$  es una función decreciente. Además,  $\varsigma(\epsilon) \geq \frac{1}{\epsilon} [(1 - \epsilon)\phi(A) + \epsilon\phi(B) - \phi(A)] = \phi(B) - \phi(A)$ .  $\Phi(A; A) = 0$  se sigue de la ecuación (3.11).

El numeral 2 es consecuencia del numeral 1, al usar el acotamiento de  $\varsigma$ .

Los detalles para la igualdad (3.12) del numeral 3 se pueden consultar en Kiefer (1974).  $\square$

**Lema 3.4.2.** *Sean  $A$  y  $B \in \text{NND}(m)$ .*

1. Si  $\phi_1, \phi_2, \dots, \phi_r$  son criterios de optimalidad cóncavos, entonces,  $\phi_c = \sum_{j=1}^r \gamma_j \phi_j$  es cóncavo, con  $\gamma_j \geq 0$  y  $\sum_j \gamma_j = 1$ , y además:  $\Phi_c(A; B)$ , la derivada direccional de  $\phi_c$  en  $A$  en la dirección de  $B$ , es:

$$\Phi_c(A; B) = \sum_{j=1}^r \gamma_j \Phi_j(A; B),$$

siendo  $\Phi_j(A; B)$  la derivada direccional de  $\phi_j$  en  $A$  en la dirección de  $B$ .

2. Si  $\phi$  es diferenciable entonces  $\Phi$  es una función lineal en la segunda componente, es decir, para  $B_1, B_2, \dots$ , matrices definidas no negativas y  $\lambda_1, \lambda_2, \dots$ , números no negativos tal que  $\sum_i \lambda_i = 1$ , se tiene:

$$\sum_j \lambda_j \Phi(A; B_j) = \Phi\left(A; \sum_j \lambda_j B_j\right).$$

3. Si  $\phi$  diferenciable,  $\Phi(M(\xi); M(\eta))$  se puede representar como:

$$\Phi(M(\xi); M(\eta)) = \int D(t; M(\xi)) \eta(dt),$$

donde  $D(t; M(\xi)) = \Phi(M(\xi); M(\delta_t))$ , y  $\delta_t$  una medida de probabilidad concentrada en el punto  $t$ .

4. Se define la tasa maximal de ascenso de  $\phi$  desde  $\xi$  como:

$$\bar{D}(\xi) = \sup_{\eta \in \Xi} \Phi(M(\xi); M(\eta)),$$

si  $\phi$  es diferenciable entonces:

$$\bar{D}(\xi) = \sup_{t \in \mathcal{X}} D(t; M(\xi)). \quad (3.13)$$

### Demostración:

Para mostrar 2, observe que:

$$\begin{aligned} \Phi\left(A; \sum_j \lambda_j B_j\right) &\stackrel{\substack{=} \\ \phi \text{ diferenciable}}}{=} \text{Tr}\left(\left(\sum_j \lambda_j B_j - A\right) \nabla \phi(A)\right) \\ &= \text{Tr}\left(\left(\sum_j \lambda_j B_j - \sum_j \lambda_j A\right) \nabla \phi(A)\right) \\ &= \text{Tr}\left(\left(\sum_j \lambda_j (B_j - A)\right) \nabla \phi(A)\right) \\ &= \sum_j \text{Tr}[(\lambda_j (B_j - A)) \nabla \phi(A)] \\ &= \sum_j \lambda_j \Phi(A; B_j). \end{aligned}$$

En el caso del numeral 3, note que al ser  $\phi$  diferenciable, se puede aplicar la ecuación 3.12 y se obtiene:

$$\begin{aligned}
 \Phi(M(\xi); M(\eta)) &= \text{Tr} [(M(\eta) - M(\xi)) \nabla \phi(M(\xi))] \\
 &= \text{Tr} \left\{ \left[ \int f(t) f^T(t) \eta(dt) - M(\xi) \right] \nabla \phi(M(\xi)) \right\} \\
 &= \text{Tr} \left[ \int f(t) f^T(t) \nabla \phi(M(\xi)) \eta(dt) - \int M(\xi) \nabla \phi(M(\xi)) \eta(dt) \right] \\
 &= \text{Tr} \left\{ \int [f(t) f^T(t) - M(\xi)] \nabla \phi(M(\xi)) \eta(dt) \right\} \\
 &= \int \text{Tr} [(M(\delta_t) - M(\xi)) \nabla \phi(M(\xi))] \eta(dt) \\
 &= \int \Phi(M(\xi), M(\delta_t)) \eta(dt)
 \end{aligned}$$

obteniendo el resultado deseado.

Para mostrar la igualdad (3.13) de 4, tenga en cuenta que como  $\delta_t \in \Xi$ , entonces:

$$\begin{aligned}
 \bar{D}(\xi) &\geq \int_{\mathcal{X}} D(t; M(\xi)) \delta_t(dt) = D(t; M(\xi)), \quad \forall t \in \mathcal{X}, \\
 \bar{D}(\xi) &\geq \sup_{t \in \mathcal{X}} D(t; M(\xi));
 \end{aligned}$$

por otro lado, al aplicar 3, se tiene:

$$\begin{aligned}
 \Phi(M(\xi); M(\eta)) &= \int D(t; M(\xi)) \eta(dt) \leq \int \sup_{t \in \mathcal{X}} D(t; M(\xi)) \eta(dt), \\
 \Phi(M(\xi); M(\eta)) &\leq \sup_{t \in \mathcal{X}} D(t; M(\xi)), \quad \forall \eta \in \Xi, \\
 \sup_{\eta \in \Xi} \Phi(M(\xi); M(\eta)) &\leq \sup_{t \in \mathcal{X}} D(t; M(\xi)).
 \end{aligned}$$

□

A continuación se enuncia y demuestra el teorema general de equivalencia, Whittle (1973) página 125.

**Teorema 3.4.1.** *Teorema general de equivalencia. Sean  $\chi$  un conjunto compacto y las funciones  $f(\cdot)$  continuas.*

1. Si  $\phi$  es un criterio de optimalidad cóncavo, entonces un diseño  $\phi$ -óptimo  $\xi^*$  es equivalentemente caracterizado por cualesquiera de las siguientes condiciones:

- a)  $\xi^*$  maximiza  $\phi$ ,
- b)  $\xi^*$  minimiza  $\bar{D}(\xi)$ ,
- c)  $\bar{D}(\xi^*) = 0$ .

2. El punto  $(M(\xi^*); M(\xi^*))$ , es un punto de silla de  $\Phi(\cdot; \cdot)$ , es decir,

$$\Phi(M(\xi^*); M(\eta)) \leq 0 = \Phi(M(\xi^*); M(\xi^*)) \leq \Phi(M(\xi); M(\xi^*)), \quad \forall \xi, \eta \in \Xi.$$

3. Si  $\phi$  es diferenciable entonces el soporte de  $\xi^*$  está contenido en

$$\{t : D(t; M(\xi^*)) = 0\},$$

y  $D(t; M(\xi^*)) = 0$ ,  $\xi^*$ -casi seguramente.

Observe que para los dos primeros incisos no se requiere que  $\phi$  sea diferenciable.

### Demostración:

De la compacidad de  $\chi$  y la continuidad de  $f(\cdot)$ , se concluye que el conjunto de todas las matrices de momentos,  $M(\Xi)$ , es compacto. Lo anterior se sigue del hecho que  $M(\Xi)$  es generado por  $S = T(\chi)$ , donde  $T$  es una aplicación continua de  $\chi$  en  $NND(m)$ , tal que  $T(t) = f(t)f^T(t)$ , y como  $\chi$  es compacto, se sigue que  $S$  es compacto, y por tanto  $M(\Xi)$  es compacto. Además, al ser  $\phi$  una función cóncava se sigue que el óptimo se alcanza en algún diseño  $\xi^* \in \Xi$ .

1a  $\Rightarrow$  1c. En efecto, de la definición de  $\Phi$  y siendo  $\xi^*$  un diseño óptimo, se tiene,

$$\Phi(M(\xi^*); M(\eta)) \leq 0, \quad \forall \eta \in \Xi,$$

de donde,

$$\bar{D}(\xi^*) \leq 0,$$

y como la igualdad se alcanza cuando  $\eta = \xi^*$ , entonces se concluye que  $\bar{D}(\xi^*) = 0$ .

1b  $\Leftrightarrow$  1c. Esta equivalencia se sigue del hecho que  $\bar{D}(\xi) \geq \Phi(M(\xi); M(\xi)) = 0$ , y se ha mostrado que la igualdad siempre se alcanza al menos cuando  $\xi$  es  $\phi$ -óptimo. Por otro lado, 1c implica 1a (y también 1b), ya que si  $\xi$  es no óptimo entonces  $\phi(M(\eta)) > \phi(M(\xi))$ , para algún  $\eta \in \Xi$ , pero:

$$\Phi(M(\xi); M(\eta)) \geq \phi(M(\eta)) - \phi(M(\xi)), \quad (3.14)$$



luego  $\Phi(M(\xi); M(\eta)) > 0$ , de donde:

$$\bar{D}(\xi) > 0.$$

Luego se ha mostrado 1.

El lado izquierdo de la expresión en 2 es evidente, pues  $\xi^*$  es  $\phi$ -óptimo. Mientras el lado derecho se sigue al tomar  $\eta = \xi^*$  en la ecuación 3.14.

Para finalizar, la afirmación 3 se sigue del par de relaciones:

$$D(t; M(\xi^*)) = \Phi(M(\xi^*); M(\delta_t)) \leq 0 \text{ y } \int D(t; M(\xi^*)) \xi^*(dt) = \Phi(M(\xi^*); M(\xi^*)) = 0. \quad \square$$

En la práctica se usa el siguiente resultado cuya prueba es sencilla usando el teorema anterior.

**Corolario 3.4.1.** *Sea  $\phi$  un criterio de optimalidad cóncavo y diferenciable. Entonces  $\xi \in \Xi$  es  $\phi$ -óptimo si y sólo si:*

$$\forall t \in \chi \quad D(t; M(\xi)) \leq 0 \text{ y } D(t; M(\xi)) = 0 \quad \forall t \in \text{Supp}(\xi). \quad (3.15)$$

## 3.5. Consecuencias del teorema general de equivalencia

A partir del Teorema general de equivalencia se tiene una forma equivalente para hallar diseños  $\phi$ -óptimos. Por el corolario 3.4.1 es suficiente encontrar la derivada direccional del criterio respectivo y hallar  $D(t; M(\xi))$ . En los siguientes corolarios se dan las expresiones para los diferentes criterios que se van a utilizar,  $\mathbb{D}$ ,  $\mathbb{D}_A$ ,  $\mathbb{L}$ , y  $\mathbb{D}_s$ -optimalidad.

**Corolario 3.5.1.**  *$\mathbb{D}$ -optimalidad, Kiefer y Wolfowitz (1960).*

*Sea  $M_{m \times m}(\xi) > 0$  y considere  $\phi(\xi) = \log |M(\xi)|$ , entonces:*

1.  $\nabla \phi(M(\xi)) = M^{-1}(\xi)$ ,
2. La derivada direccional de  $\phi$  en  $\xi$  en la dirección del diseño  $\eta$  es:

$$\Phi(M(\xi); M(\eta)) = \text{Tr}(M(\eta) M^{-1}(\xi)) - m,$$

3.  $D(t; M(\xi)) = \text{Tr}(f(t) f^T(t) M^{-1}(\xi)) - m = f^T(t) M^{-1}(\xi) f(t) - m$

**Demostración:**

Para ver 2 note que:

$$\begin{aligned}\Phi(M(\xi); M(\eta)) &= \Phi(M(\xi); M(\eta)) = \text{Tr}((M(\eta) - M(\xi)) \nabla \phi(M(\xi))) \\ &= \text{Tr}((M(\eta) - M(\xi)) M^{-1}(\xi)) \\ &= \text{Tr}(M(\eta) M^{-1}(\xi)) - m\end{aligned}$$

Mientras 3 es un caso particular de 2, en efecto:

$$\begin{aligned}D(t; M(\xi)) &= \Phi(M(\xi); M(\delta_t)) = \Phi(M(\xi); f(t) f^T(t)) \\ &= \text{Tr}(f(t) f^T(t) M^{-1}(\xi)) - m = f^T(t) M^{-1}(\xi) f(t) - m.\end{aligned}$$

□

**Corolario 3.5.2.**  $\mathbb{L}$ -*optimalidad*.

Si  $\phi(M(\xi)) = -\text{Tr}(K^T M^{-1}(\xi) K) = -\text{Tr}(LM^{-1}(\xi))$ ,  $L = KK^T$ . Entonces:

1.  $\nabla \phi(M(\xi)) = M^{-1}(\xi) LM^{-1}(\xi)$ .
2. La derivada direccional de  $\phi$  en  $\xi$  en la dirección del diseño  $\eta$  está dada por:

$$\Phi(M(\xi); M(\eta)) = \text{Tr}((M(\eta) - M(\xi)) M^{-1}(\xi) LM^{-1}(\xi)).$$

3.  $D(t; M(\xi)) = f^T(t) M^{-1}(\xi) LM^{-1}(\xi) f(t) - \text{Tr}(LM^{-1}(\xi))$ .

**Demostración:**

La parte 2 es directa,

$$\Phi(\xi; \eta) = \Phi(M(\xi); M(\eta)) = \text{Tr}((M(\eta) - M(\xi)) M^{-1}(\xi) LM^{-1}(\xi))$$

La parte 3 se sigue de:

$$\begin{aligned}D(t; M(\xi)) &= \Phi(M(\xi); M(\delta_x)) = \text{Tr}((f(t) f^T(t) - M(\xi)) M^{-1}(\xi) LM^{-1}(\xi)) \\ &= f^T(t) M^{-1}(\xi) LM^{-1}(\xi) f(t) - \text{Tr}(LM^{-1}(\xi)).\end{aligned}$$

□

**Corolario 3.5.3.**  $\mathbb{D}_A$ -optimalidad,  $\mathbb{D}$ -optimalidad para estimar el sistema  $A^T\theta$ .  
 Sea  $A_{m \times m'}$ , con  $r(A) = m'$ . Defina  $\phi(M(\xi)) = -\log |A^T M^{-1}(\xi) A|$ . Entonces:

1.  $\frac{\partial \log |A^T M^{-1}(\xi) A|}{\partial M} = -M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{-1} A^T M^{-1}(\xi)$ .

2. La derivada direccional de  $\phi$  en  $\xi$  en la dirección del diseño  $\eta$  está dada por:

$$\Phi(M(\xi); M(\eta)) = \text{Tr} \left[ (M(\eta) - M(\xi)) M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{-1} A^T M^{-1}(\xi) \right].$$

3.  $D(t; M(\xi)) = f^T(t) M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{-1} A^T M^{-1}(\xi) f(t) - m'$ .

**Demostración:**

Para mostrar 1 se usa el hecho de que:  $\log |U| = \lim_{p \rightarrow 0} p^{-1} (\text{Tr} U^p - m)$ . Se tiene entonces que:

$$\begin{aligned} \log |A^T M^{-1}(\xi) A| &= \lim_{p \rightarrow 0} p^{-1} (\text{Tr} (A^T M^{-1}(\xi) A)^p - m) \\ \frac{\partial \log |A^T M^{-1}(\xi) A|}{\partial M} &= \lim_{p \rightarrow 0} p^{-1} \frac{\partial}{\partial M(\xi)} [\text{Tr} (A^T M^{-1}(\xi) A)^p - m] \\ &\stackrel{\text{ver Kiefer (1974)}}{=} \lim_{p \rightarrow 0} -M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{p-1} A^T M^{-1}(\xi) \\ &= -M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{-1} A^T M^{-1}(\xi) \end{aligned}$$

Para ver 2 se tiene:

$$\begin{aligned} \Phi(M(\xi); M(\eta)) &= -\text{Tr} \left[ (M(\eta) - M(\xi)) \left( -M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{-1} A^T M^{-1}(\xi) \right) \right] \\ &= \text{Tr} \left[ (M(\eta) - M(\xi)) M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{-1} A^T M^{-1}(\xi) \right] \end{aligned}$$

Y por último se muestra 3:

$$\begin{aligned} D(t; M(\xi)) &= D(M(\xi); M(\delta_t)) \\ &= f^T(t) M^{-1} A (A^T M^{-1} A)^{-1} A^T M^{-1} f(t) - \text{Tr} \left( A (A^T M^{-1} A)^{-1} A^T M^{-1} \right) \\ &= f^T(t) M^{-1}(\xi) A (A^T M^{-1}(\xi) A)^{-1} A^T M^{-1}(\xi) f(t) - m'. \end{aligned}$$

siendo  $m'$  el número de filas de  $A$ . □

En el siguiente corolario se presenta el criterio usado para estimar en forma óptima un subconjunto de  $v$ -parámetros,  $v < m$ . Además se puede usar para hacer discriminación entre dos modelos anidados.

**Corolario 3.5.4.** *D<sub>v</sub>-optimalidad.*

Sea  $\phi_v(M(\xi)) = \log \frac{|M(\xi)|}{|M_{22}(\xi)|}$ , donde  $M(\xi) = \begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) \end{bmatrix}$ , con  $M_{22}(\xi)$  matriz de orden  $l \times l$ , y  $f^T(t)\beta = f_1^T(t)\beta_1 + f_2^T(t)\beta_2$ , y al reescribir  $\phi_v$  como:

$$\phi_v(M(\xi)) = \log |M(\xi)| - \log |M_{22}(\xi)| = \phi_1(M(\xi)) - \phi_2(M(\xi)),$$

entonces:

1.  $\Phi(M(\xi); M(\eta)) = \text{Tr}[(M(\eta) - M(\xi))M^{-1}(\xi)] - \text{Tr}[(M_{22}(\eta) - M_{22}(\xi))M_{22}^{-1}(\xi)]$
2.  $D(t; M(\xi)) = f^T(t)M^{-1}(\xi)f(t) - f_2^T(t)M_{22}^{-1}(\xi)f_2(t) - v$ , donde  $v = m - l$ .

**Demostración:**

Para mostrar 1 se observa que:

$$\begin{aligned} \Phi(M(\xi); M(\eta)) &= \Phi_1(M(\xi), M(\eta)) - \Phi_2(M(\xi), M(\eta)) \\ &= \text{Tr}[(M(\eta) - M(\xi))M^{-1}(\xi)] - \text{Tr}[(M_{22}(\eta) - M_{22}(\xi))M_{22}^{-1}(\xi)] \end{aligned}$$

A partir de la expresión anterior y usando propiedades de la derivada direccional se obtiene 2, en efecto:

$$\begin{aligned} D(t; M(\xi)) &= f^T(t)M^{-1}(\xi)f(t) - m - f_2^T(t)M_{22}^{-1}(\xi)f_2(t) + l \\ &= f^T(t)M^{-1}(\xi)f(t) - f_2^T(t)M_{22}^{-1}(\xi)f_2(t) - v. \end{aligned}$$

□

### 3.6. Diseños óptimos para modelos no lineales

Los modelos no lineales se pueden representar por:

$$Y(x) = \eta(x; \theta) + \epsilon, \quad (3.16)$$

donde, como en el modelo lineal, las variables explicativas  $x^T = (x_1, x_2, \dots, x_k)$  varían en un espacio de diseño compacto,  $\chi \subseteq \mathbb{R}^k$ , dotado de una  $\sigma$ -álgebra,  $\mathcal{B}$ , (Borelianos en  $\chi$ , agregándole los conjuntos unitarios),  $\theta \in \Omega \subseteq \mathbb{R}^m$ , los errores con media cero y varianza constante y  $\eta(x; \theta)$  es una función no lineal en  $\theta$ .

En el modelo 3.16, dado un diseño  $\xi$  definido en  $\mathcal{B}$ , se conoce que el estimador de mínimos cuadrados para  $\theta$ ,  $\hat{\theta}_n$ , bajo ciertas condiciones de regularidad (entre otras, como continuidad de  $\eta(x; \theta)$  con respecto a  $\theta \in \Omega$ , para cualquier  $x$ ) es asintóticamente insesgado, de hecho es fuertemente consistente, es decir,  $\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$  casi seguramente. Además, la expresión de la matriz de varianzas-covarianzas de  $\hat{\theta}$  se halla al aproximar  $\eta(x; \theta)$  por una expansión en potencias en una vecindad de  $\theta_0$ :

$$\eta(x; \theta) \cong \eta(x; \theta_0) + (\theta - \theta_0)^T f(x; \theta_0),$$

siendo  $f(x; \theta) = \frac{\partial \eta(x; \theta)}{\partial \theta}$ .

Se puede mostrar, ver por ejemplo Fedorov y Hackl (1997), que la matriz de varianzas-covarianzas asintótica de  $\hat{\theta}$  es la inversa de la matriz:

$$M(\xi; \theta) = \mathbb{E}_\xi [f(x; \theta) f^T(x; \theta)] = \int_{\mathcal{X}} f(x; \theta) f^T(x; \theta) d\xi(x),$$

de hecho bajo algunos supuestos, como por ejemplo la existencia de  $\partial^2 \eta(x; \theta) / \partial \theta_i \partial \theta_j$ ,  $i, j = 1, 2, \dots, m$ , se puede mostrar que  $\sigma^{-2} n D[\hat{\theta}_n] \rightarrow M^{-1}(\theta_0)$ , donde  $M(\theta_0) = \lim_{n \rightarrow \infty} \sigma^2 n^{-1} M(\xi_{(n)}, \theta_0)$ , donde  $\xi_{(n)}$  es un diseño exacto. Es decir,  $\sigma^{-2} n M^{-1}(\xi_{(n)}, \hat{\theta})$  es un estimador fuertemente consistente para  $M^{-1}(\theta_0)$ .

De esta forma se motiva el estudio de funcionales de la matriz  $M(\xi; \theta)$ , debido a sus propiedades asintóticas. Note que bajo el supuesto de normalidad e independencia en los errores se tiene que  $M(\xi; \theta)$  es una combinación ponderada de las matrices de información de Fisher asociada a cada punto de soporte  $x_u$  del diseño  $\xi$ ,  $u = 1, 2, \dots, s$ , esta información está dada por:

$$\mathbb{E} \left[ \frac{\partial}{\partial \theta} \log p(y|x_u, \theta) \left( \frac{\partial}{\partial \theta} \log p(y|x_u, \theta) \right)^T \right] = \mathbb{E} \left[ -\frac{\partial^2 \log p(y|x_u, \theta)}{\partial \theta^2} \right],$$

donde  $p(y|x_u, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\eta(x_u; \theta))^2}{2\sigma^2}\right)$ . Luego la componente  $(i, j)$  de la matriz de información de Fisher del punto  $x_u$  se reduce a:

$$\begin{aligned} \mathbb{E} \left[ -\frac{\partial^2 \left( -\frac{(y-\eta(x_u; \theta))^2}{2\sigma^2} \right)}{\partial \theta_i \partial \theta_j} \right] &= \mathbb{E} \left[ \frac{1}{\sigma^2} \left( -y \frac{\partial^2 \eta(x_u; \theta)}{\partial \theta_i \partial \theta_j} + \eta(x_u; \theta) \frac{\partial^2 \eta(x_u; \theta)}{\partial \theta_i \partial \theta_j} + \frac{\partial \eta(x_u; \theta)}{\partial \theta_i} \frac{\partial \eta(x_u; \theta)}{\partial \theta_j} \right) \right] \\ &= \frac{1}{\sigma^2} \frac{\partial \eta(x_u; \theta)}{\partial \theta_i} \frac{\partial \eta(x_u; \theta)}{\partial \theta_j}. \end{aligned}$$

Luego la matriz de información de Fisher del punto  $x_u$  está dada por:  $\frac{1}{\sigma^2} f(x_u; \theta) f^T(x_u; \theta)$ , y así,

$$M(\xi; \theta) = \sum_{u=1}^s w_u \frac{1}{\sigma^2} f(x_u; \theta) f^T(x_u; \theta).$$

Sin pérdida de generalidad se supone que  $\sigma^2 = 1$ , ya que en los criterios de optimalidad que se van a usar no se ven afectados por el valor de este parámetro, se supone que este parámetro es constante. De lo anterior a  $M(\xi; \theta)$  se le conoce como matriz de información, y juega el papel de la matriz de momentos del modelo lineal, si se considerara el modelo linealizado.

La dependencia de  $M$  de  $\theta$  hace que la búsqueda de diseños óptimos dependa de este parámetro. En forma análoga al caso lineal, se cuantifica la magnitud de la información suministrada por  $M(\xi, \theta)$  a partir de funcionales de ésta, y consecuentemente la maximización de alguna función de información,  $\phi$ , de valor real, ver Box y Lucas (1959). En este tipo de modelos las propiedades que tienen los criterios de optimalidad se traducen en términos asintóticos.

Para la construcción de los diseños óptimos existen varias alternativas, en este trabajo se exploran las siguientes:

### 3.6.1. Diseños óptimos locales

Introducidos por Chernoff (1953), son los primeros diseños que aparecieron para el caso no lineal. Consiste en dar inicialmente un valor a priori para  $\theta$ ,  $\theta_0$ , que esté cercano al valor verdadero del parámetro. Esta asignación a priori dependerá del conocimiento que tenga el investigador acerca del problema, de donde se puede optar por estimaciones de  $\theta$  a partir de datos previos, o a partir de estudios similares. Una vez se tenga el valor a priori para  $\theta$ , denotado por  $\theta_0$ , se construyen diseños óptimos a partir de funcionales de la matriz de información  $M(\xi; \theta_0)$ . Como lo afirma Cochran (1973) página 771: “Denme un valor de  $\theta$  y aseguro diseñar el mejor experimento para estimar  $\theta$ ”. Es claro que los diseños óptimos en modelos no lineales dependerán del valor a priori que se halla seleccionado para  $\theta$ , por tanto es importante tener un buen estimador inicial de  $\theta$ ,  $\theta_0$ . Los diseños resultantes son diseños óptimos locales. Varios autores han usado esta alternativa para la construcción de los diseños, ver por ejemplo: Ford, et al. (1992); Dette, Melas y Pepelyshev (2004b); Dette, Melas y Wong (2005); O’Brien (1992); O’Brien (2005), entre otros. La construcción de diseños  $\mathbb{D}$ -óptimos locales y  $\mathbb{L}$ -óptimos locales, son explorados en los ejemplos de la subsección 3.6.3.

### 3.6.2. Diseños óptimos promediados por una distribución a priori $\pi$

Los diseños óptimos locales dependen de una buena elección para  $\theta$ , aunado al hecho de que en varios modelos los diseños óptimos resultantes son diseños no aptos para la realización de pruebas de bondad de ajuste, ver los ejemplos de la subsección 3.6.3. Una posible solución se presenta cuando se conoce una serie de posibles valores para  $\theta$  con una distribución asociada. En general, si se tiene una distribución a priori para  $\theta$ ,  $\pi$ , se puede optar por usar como criterio de optimalidad el resultado de integrar el criterio subyacente con respecto a la medida dada por  $\pi$ . Dando como resultado el criterio de optimalidad, que en algunos artículos se le denomina, Bayesiano. Se hace la anotación que en este trabajo se usa el término Bayesiano por su relación con el uso de distribuciones a priori. Es claro que los diseños óptimos propiamente Bayesianos, hacen uso de funciones de utilidad, ver por ejemplo la revisión de Chaloner y Verdinelli (1995) y juegan un papel importante en la teoría de los diseños óptimos. Cabe señalar que en este trabajo no se hizo uso de este enfoque.

En algunas situaciones considerando distribuciones a prioris lo suficientemente dispersas se pueden encontrar diseños óptimos con un número suficiente de puntos de soporte donde es posible llevar a cabo pruebas de bondad de ajuste, ver Atkinson y Donev (1992). En particular, un diseño  $\xi$  es  $\mathbb{D}$ -óptimo Bayesiano (con respecto a la distribución a priori  $\pi$ ), para abreviar  $\mathbb{D}_\pi$ -óptimo, si maximiza:

$$\mathbb{E}_\pi [\log |M(\xi; \theta)|] = \int_{\Omega} \log |M(\xi; \theta)| d\pi(\theta).$$

En general, un diseño es  $\phi$ -óptimo Bayesiano con respecto a la distribución a priori  $\pi$ , abreviado por  $\phi_\pi$ -óptimo, Dette, Haines e Imhof (2007), si maximiza:  $E_\pi [\phi(M(\xi; \theta))]$ . Ejemplos de este tipo de diseños son mostrados en la subsección 3.6.3 y también en el capítulo 4 donde se exploran otros criterios de optimalidad.

Para  $\mathbb{D}_\pi$  optimalidad, se obtiene la siguiente equivalencia, generalización del teorema de Kiefer y Wolfowitz, y se pueden usar los resultados dados en la sección 3.4, bajo el supuesto que  $M > 0$ :

$$\xi \text{ es } \mathbb{D}_\pi - \text{óptimo si y sólo sí } \mathbb{E} [f^T(x; \theta)M^{-1}(\xi; \theta)f(x; \theta)] \leq m, \quad \forall x \in \chi. \quad (3.17)$$

La respectiva equivalencia se obtiene para  $\mathbb{L}_\pi$ -optimalidad al calcular la esperanza, con respecto a la a priori  $\pi$ , de la expresión 3.9. Es decir,  $\xi$  es  $\mathbb{L}_\pi$ -óptimo si y sólo sí:

$$\forall x \in \chi, \quad \mathbb{E}_\pi [f^T(x; \theta)M^{-1}(\xi)KK^TM^{-1}f(x; \theta) - \text{Tr}(K^TM^{-1}K)] \leq 0, \quad (3.18)$$

donde  $K$  y  $M^{-1}$ , son matrices que dependen de  $\theta$ .

En la literatura de diseños óptimos en modelos no lineales, existen otras alternativas de solución que no son exploradas en este trabajo, como lo son los diseños maximín, y maximín estandarizados, entre otros, ver Dette, Haines e Imhof (2007), Dette y Sahn (1998).

A continuación se presentan varios ejemplos de modelos no lineales donde el interés está en encontrar diseños óptimos tanto óptimos locales como promediados por una distribución a priori: diseños óptimos Bayesiano. Estos ejemplos aparecen en la literatura usual de diseños óptimos, y en este trabajo se han encontrado los diseños óptimos para otros valores locales y considerando distribuciones a priori diferentes.

### 3.6.3. Ejemplos de diseños óptimos en modelos no lineales

#### Ejemplo 3.6.1. Modelo de decaimiento exponencial.

*Este modelo está dado por:*

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

*Si  $\theta_0$  es una buena asignación para  $\theta$ , su matriz de información, que en efecto es un número real, 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  $\mathbb{D}$ -óptimo local concentra toda su masa en un punto. Se verá a continuación que el punto es:  $x_0 = 1/\theta_0$ . Sea  $\xi_0$  el diseño que tiene como punto de soporte a  $x_0$ , entonces:*

$$M(\xi_0; \theta_0) = x_0^2 \exp(-2\theta_0 x_0), \quad (3.19)$$

*No es difícil mostrar que el máximo de la ecuación 3.19 se alcanza en  $x_0 = 1/\theta_0$ , y*

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

*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  $\mathbb{D}$ -óptimo local. Este diseño no permite realizar pruebas*



de bondad de ajuste para el modelo en cuestión. El diseño depende de la especificación de  $\theta_0$ , y puede llegar a ser ineficiente si  $\theta_0$  está muy lejos del valor verdadero  $\theta$ .

Otra forma de hallar un diseño óptimo es a partir de un enfoque Bayesiano, donde se incorpora el conocimiento acerca de  $\theta$  por medio de una distribución a priori. Como ilustración se consideran 6 distribuciones a prioris discretas, uniformes en 5 puntos, y se hallaron los respectivos diseños  $\mathbb{D}_\pi$ -óptimos para estimar  $\theta$ , con las diferentes a priori, ver Tabla 3.1. Lo anterior fue hecho numéricamente con ayuda de algoritmos computacionales programados en el lenguaje R usando la equivalencia 3.17. Los diseños  $\mathbb{D}_\pi$ -óptimos obtenidos están formados por tres puntos de soporte, observándose variación en las distintas a prioris consideradas, tanto en los puntos de soporte (primera fila de cada diseño) como en sus pesos (segunda fila del diseño). En Atkinson y Donev (1992), página 230, muestran como los puntos de soporte del diseño aumentan a medida que la distribución a priori que los autores consideran es más dispersa.

$\pi$	Soporte dist. a priori $\pi_j$	Diseño $\mathbb{D}_{\pi_j}$ -óptimo
$\pi_1$	(0.09, 0.49, 1, 4.9, 9)	0.16 1.50 10.99 0.44 0.40 0.16
$\pi_2$	(0.10 0.50, 1, 5.0, 10)	0.14 1.52 9.81 0.43 0.42 0.15
$\pi_3$	(0.11 0.51, 1, 5.1, 11)	0.13 1.54 8.81 0.43 0.44 0.13
$\pi_4$	(0.12 0.52, 1, 5.2, 12)	0.12 1.56 7.95 0.42 0.46 0.12
$\pi_5$	(0.14 0.54, 1, 5.4, 14)	0.11 1.62 6.55 0.42 0.50 0.08
$\pi_6$	(0.15 0.55, 1, 5.5, 15)	0.10 1.65 5.96 0.42 0.52 0.06

Tabla 3.1: Diseños óptimos para el modelo de decaimiento exponencial con diferentes distribuciones a priori uniformes para  $\theta$ .

### Ejemplo 3.6.2. Modelo de compartimientos.

Los modelos de compartimientos son de gran utilidad en farmacocinética, ver capítulo 2. Estos modelos se utilizan, 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:

$$\eta(x, \theta) = \frac{\theta_1}{\theta_1 - \theta_2} \{ \exp(-\theta_2 x) - \exp(-\theta_1 x) \} \quad x \geq 0, \theta_1 > \theta_2 > 0. \quad (3.20)$$

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

1. El área bajo la curva (AUC):  $g_1(\theta) = \int_0^\infty \eta(x; \theta) dx = \frac{1}{\theta_2}$ .
2. Tiempo para la concentración máxima:  $g_2(\theta) = x_{\text{máx}} = \frac{\log \theta_1 - \log \theta_2}{\theta_1 - \theta_2}$ .
3. La concentración máxima:  $g_3(\theta) = \eta(x_{\text{máx}}; \theta)$ .

La construcción de diseños óptimos para la estimación de estas funciones simultáneamente, se puede hacer a partir de diseños  $\mathbb{L}$ -óptimos locales (ver ecuación 3.7, con  $p = -1$ ), y diseños  $\mathbb{L}$ -óptimos promediados por una distribución a priori, se consideró una a priori uniforme. La  $j$ -ésima columna,  $K_j$ , de  $K$  es el gradiente de la función no lineal,  $g_j(\theta)$ , evaluada en  $\theta_0^T = (\theta_{10}, \theta_{20})$ . Así se asegura que el diseño óptimo será aquel que minimice el promedio asintótico de la 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:

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

en forma análoga se halla la tercera columna de  $K$ . Como ilustración, se tomó  $\theta_0^T = (0.7, 0.2)$ , y en la Tabla 3.2 se presentan los diseños  $\mathbb{L}$ -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,  $\theta$ :

$$\Theta = \left\{ (0.70, 0.20)^T, (0.65, 0.15)^T, (0.75, 0.25)^T, (0.65, 0.25)^T, (0.75, 0.15)^T \right\},$$

es decir,  $\pi(\theta) = 1/5$ ,  $\forall \theta \in \Theta$ . En la Tabla 3.2 se reporta el diseño  $\mathbb{L}$ -óptimo Bayesiano obtenido. Ambos diseños  $\mathbb{L}$ -óptimo local y  $\mathbb{L}$ -óptimo promediado por la a priori  $\pi$ , presentan pocas diferencias. Además, en ambos casos se verifica que el diseño hallado satisface las equivalencias dadas por las ecuaciones 3.9 y 3.18, respectivamente.

Criterio	Diseño óptimo	
$\mathbb{L}$ -optimalidad local	1.31 6.60	
	0.28 0.72	
$\mathbb{L}$ -optimalidad Promediada por la a priori $\pi$	1.46 7.14	
	0.27 0.73	

Tabla 3.2: Diseños  $\mathbb{L}$ -óptimos locales y promediados por la a priori  $\pi$  para el modelo 3.20

Hay muchos otros trabajos en diseños de experimentos óptimos para modelos de compartimientos. Por ejemplo, trabajos recientes de López y Ramos (2007c) y López y Ramos (2007b). En el primer trabajo se estudia otro modelo de cuatro compartimientos con posible aplicación en farmacocinética y se hallan diseños  $\mathbb{L}$ -óptimos locales Bayesianos y en el segundo trabajo, usando el mismo modelo, se hacen comparaciones entre diseños  $\mathbb{L}$  y  $\mathbb{D}_A$  óptimos Bayesianos. Para otros trabajos en modelos de compartimientos con observaciones correlacionadas ver por ejemplo López-Fidalgo et al (2005).

### 3.7. Criterios de optimalidad para discriminar entre modelos

En los contextos tanto lineal como no lineal se asume implícitamente que se conoce la forma funcional  $\eta(x; \theta)$ , lo cual en situaciones prácticas no es del todo cierto. Es posible tener idea de que uno de varios modelos es el adecuado para la situación experimental que se está estudiando, pero no se sabe cual es el modelo correcto. Lo anterior hace pensar en la necesidad de construir diseños que tengan una potencia alta para discriminar entre varios modelos, es decir que permitan decidir cual de los modelos en cuestión es el más adecuado.

En lo que sigue se supone que todos los modelos bajo consideración dependen del mismo conjunto de variables independientes. Se presenta el caso donde se necesita discriminar entre dos modelos. Se supone que los dos modelos tienen la siguiente estructura:

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

donde la función  $\eta(x; \theta)$  es una de dos funciones conocidas  $\eta_1(x; \theta_1)$  ó  $\eta_2(x; \theta_2)$ , siendo  $\theta_1 \in \Omega_1 \subset \mathbb{R}^{m_1}$  y  $\theta_2 \in \Omega_2 \subset \mathbb{R}^{m_2}$ .

A continuación se presenta un posible criterio para discriminar entre dos modelos competitivos. Es uno de los criterios más usados en la literatura de los diseños óptimos, en el Capítulo 4, sección 4.3.6, donde se compara el criterio que se verá a continuación con el criterio  $\mathbb{D}_s$ -optimal para el caso de dos modelos anidados.

### 3.7.1. $\mathbb{T}$ -optimalidad

Este criterio fue introducido por Atkinson y Cox (1974), y luego fue generalizado para el caso de discriminación entre más de dos modelos por Atkinson y Fedorov (1975b). Los resultados que los autores obtienen dependen de los estimadores de mínimos cuadrados y del supuesto de normalidad. Para el caso de no normalidad se ha propuesto un criterio alternativo ver López-Fidalgo et al (2007).

El diseño óptimo para discriminar depende de cual de las funciones,  $\eta_i$ , es la correcta. Suponga que la función correcta es  $\eta_v(x) = \eta(x; \theta) = \eta_1(x; \theta_1)$ . El criterio de optimalidad está basado en la siguiente idea: Buscar un buen diseño para discriminar entre  $\eta_v$  y  $\eta_2$  de tal forma que suministre una suma de cuadrados de falta de ajuste grande para el segundo modelo. Si el segundo modelo se ajusta a los datos, los estimadores de mínimos cuadrados de los parámetros dependerán del diseño experimental, del valor de  $\theta_1$  y la distribución de los errores. Los parámetros estimados se hallan a partir de la siguiente expresión:

$$\begin{aligned}\hat{\theta}_2(\xi) &= \arg \min_{\theta_2} \int_{\mathcal{X}} \{\eta_v(x) - \eta_2(x; \theta_2)\}^2 d\xi(x) \\ &= \arg \min_{\theta_2} \sum_{x \in \text{Supp}(\xi)} \{\eta_v(x) - \eta_2(x; \theta_2)\}^2 \xi(x)\end{aligned}$$

y este valor produce la siguiente suma de cuadrados residual:

$$\Delta_2(\xi) = \min_{\theta_2} \Delta_{21}(\xi; \theta_2). \quad (3.21)$$

donde:

$$\Delta_{21}(\xi; \theta_2) = \int_{\mathcal{X}} \{\eta_v(x) - \eta_2(x; \theta_2(\xi))\}^2 d\xi(x).$$

$\Delta_2(\xi)$  es una medida de falta de ajuste cuando  $\eta_2(x; \hat{\theta}_2)$  se reemplaza por  $\eta(x)$ . En el caso donde  $\Omega_2 = \mathbb{R}^{m_2}$ ,  $\eta_2(x; \theta_2) = \theta_2^T f_2(x)$ , y los errores son normales e independientes, se tiene que  $\Delta_2(\xi)$  es proporcional al parámetro de no centralidad de la distribución  $\chi^2$  de la suma de cuadrados residual asociada al segundo modelo. Aquellos diseños que maximizan  $\Delta_2(\xi)$  se denominan diseños  $\mathbb{T}$ -óptimos, la letra  $\mathbb{T}$  se usa con el fin de enfatizar la conexión con contraste de modelos (“Testing models”).

Para el caso de modelos lineales en los parámetros, los diseños  $\mathbb{T}$ -óptimos obtenidos al maximizar (3.21) dan la prueba  $F$  más potente para la falta de ajuste del segundo modelo cuando el primero es cierto. En el caso de modelos no lineales en los parámetros las propiedades de la prueba  $F$  exacta se reemplazan por propiedades asintóticas, pero aun así el diseño obtenido será aquel que maximice la ecuación 3.21.

Existe un teorema de equivalencia para el criterio definido en 3.21, ver Atkinson y Fedorov (1975a). Se usa el hecho de que  $\Delta_{21}(\xi, \theta_2)$  es una función lineal en  $\xi$ , es cóncava y diferenciable y bajo el supuesto de que  $\hat{\theta}_2(\xi)$  es único. Se puede mostrar que la función de sensibilidad está dada por:  $D(x; \xi) = \Psi_2(x; \xi) - \Delta_2(\xi)$ , donde  $\Psi_2(t, \xi) = \{\eta_v(x; \theta_1) - \eta_2(x; \theta_2)\}^2$  y  $\theta_2 = \hat{\theta}_2(\xi)$ .

El teorema se enuncia a continuación.

**Teorema 3.7.1.** *Con las condiciones dadas anteriormente se tiene que: Un diseño  $\xi^*$  es  $\mathbb{T}$ -óptimo si y solo si  $\Psi_2(x; \xi^*) - \Delta_2(\xi^*) \leq 0 \forall x \in \chi$  con igualdad en los puntos de soporte del diseño  $\xi^*$ , con  $\Psi_2(t, \xi^*) = \{\eta_v(x; \theta_1) - \eta_2(x; \theta_2^*)\}^2$ ,  $\theta_2^* = \hat{\theta}_2(\xi^*)$ , y  $\Delta_2(\xi^*) = \int_{\chi} \{\eta_v(x; \theta_1) - \eta_2(x; \theta_2^*)\}^2 d\xi^*(x)$ .*

Se observa que los diseños  $\mathbb{T}$ -óptimos realmente son diseños locales, pues dependen del valor de  $\theta$  para el modelo que se asume como correcto. Si se cuenta con una distribución a priori,  $\pi$ , para  $\theta$ , entonces se puede hablar de diseños  $\mathbb{T}_{\pi}$ -óptimos promediados con respecto a la distribución a priori  $\pi$ .

En la siguiente subsección se presentan varios ejemplos, que aparecen en la literatura, donde se construyen diseños óptimos usando otras distribuciones a prioris.

### 3.7.2. Algunos ejemplos de diseños $\mathbb{T}$ -óptimos

A continuación se van a explorar algunos ejemplos que aparecen en la literatura de diseños óptimos para discriminación entre dos modelos, ver Atkinson y Donev (1992). Adicional a los diseños allí reportados se encuentran otros diseños.

**Ejemplo 3.7.1.** *Discriminación entre los siguientes dos modelos:*

$$\eta_1(x; \theta_1) = \theta_{10} + \theta_{11}e^x + \theta_{12}e^{-x} \text{ vs. } \eta_2(x; \theta_2) = \theta_{20} + \theta_{21}x + \theta_{22}x^2,$$

para  $-1 \leq x \leq 1$ .

Se considera el primer modelo como verdadero. Entonces el diseño  $\mathbb{T}$ -óptimo depende de los valores de los parámetros  $\theta_{11}$  y  $\theta_{12}$ . Por ejemplo Atkinson y Fedorov (1975a) consideran el vector de parámetros:  $(4.5, -1.5, -2)^T$ , de donde:

$$\eta_v(x) = \eta_1(x; (4.5, -1.5, -2)^T) = 4.5 - 1.5e^x - 2e^{-x}.$$

Los autores justifican la aproximación de la función anterior por un polinomio cuadrático, al observar que en  $x = -1$ ,  $\eta_v(-1) = -1.488$ , y alcanza su máximo de 1.036 en  $x = 0.144$  antes de declinarse a  $-0.131$  en  $x = 1$ .

$\theta_{11}$	$\theta_{12}$	Diseño óptimo $\xi^*$	$\tilde{\theta}_2^T$	$\Delta_2(\xi^*)$
-1.0	-2.5	$\begin{bmatrix} -1.000 & -0.607 & 0.339 & 1.000 \\ 0.222 & 0.390 & 0.278 & 0.109 \end{bmatrix}$	( 1.04 1.69 -1.94 )	$5.14 \times 10^{-3}$
-0.5	-3.0	$\begin{bmatrix} -1.000 & -0.577 & 0.407 & 1.000 \\ 0.204 & 0.370 & 0.296 & 0.130 \end{bmatrix}$	( 1.04 2.82 -1.94 )	$1.13 \times 10^{-2}$
0.0	-3.5	$\begin{bmatrix} -1.000 & -0.560 & 0.437 & 1.000 \\ 0.195 & 0.360 & 0.305 & 0.140 \end{bmatrix}$	( 1.04 3.95 1.94 )	$2.48 \times 10^{-2}$
0.5	-4.0	$\begin{bmatrix} -1.000 & -0.550 & 0.453 & 1.000 \\ 0.189 & 0.354 & 0.311 & 0.146 \end{bmatrix}$	( 1.04 5.09 -1.94 )	$4.05 \times 10^{-2}$
0.5	4.0	$\begin{bmatrix} -1.000 & -0.572 & 0.416 & 1.000 \\ 0.201 & 0.367 & 0.299 & 0.133 \end{bmatrix}$	( 8.95 -3.95 2.49 )	$2.53 \times 10^{-2}$

Tabla 3.3: Diseños  $\mathbb{T}$ -óptimos locales cuando  $\eta_1(x; \theta_1)$  es la función correcta y  $\theta_{10} = 4.5$ .

Se puede verificar entonces, que el diseño  $\mathbb{T}$ -óptimo para discriminar entre los dos modelos, es decir, el diseño para el cual  $\Delta_2(\xi)$  es maximizado, está dado por:

$$\xi^* = \begin{bmatrix} -1.000 & -0.669 & 0.144 & 0.957 \\ 0.253 & 0.428 & 0.247 & 0.072 \end{bmatrix},$$

donde  $\Delta_2(\xi^*) = 1.087 \times 10^{-3}$ . Un hecho sobresaliente de este diseño es que la mitad del peso se divide entre el primer y tercer punto del diseño y la otra mitad entre los otros dos.

En la Tabla 3.3 se presentan los diseños  $\mathbb{T}$  óptimos hallados para otros valores de  $\theta_{11}$  y  $\theta_{12}$ . También se reporta el valor del vector de parámetros  $\tilde{\theta}_2$  asociado al diseño óptimo, además del valor óptimo alcanzado por el respectivo diseño. Se observa que el diseño óptimo no depende del valor de  $\theta_{10}$ , sin embargo se utilizó 4.5 para los cálculos. Mediante el uso del teorema de equivalencia se verificó que en efecto todos los diseños obtenidos son  $\mathbb{T}$ -óptimos locales.

En la Tabla 3.4 se encuentran los diferentes diseños  $\mathbb{T}$ -óptimos para el caso donde  $\eta_2(x; \theta_2)$  es el modelo verdadero y diferentes valores de  $\theta_2$ . Adicionalmente se reporta el valor del vector de parámetros,  $\theta_1$ , para el diseño óptimo y el valor máximo del criterio,  $\Delta_2(\xi^*)$ . Se observa que todos los diseños óptimos locales tienen cuatro puntos de soporte, excepto el tercer diseño óptimo local que tiene cinco. En la Figura 3.3 se muestra que en efecto todos los diseños obtenidos son  $\mathbb{T}$ -óptimos locales.

Adicionalmente, con propósitos ilustrativos, se construyeron diseños  $\mathbb{T}$ -óptimos promediados por una distribución a priori. Inicialmente se considera una distribución uniforme en los cinco valores considerados del vector de parámetros para el caso donde

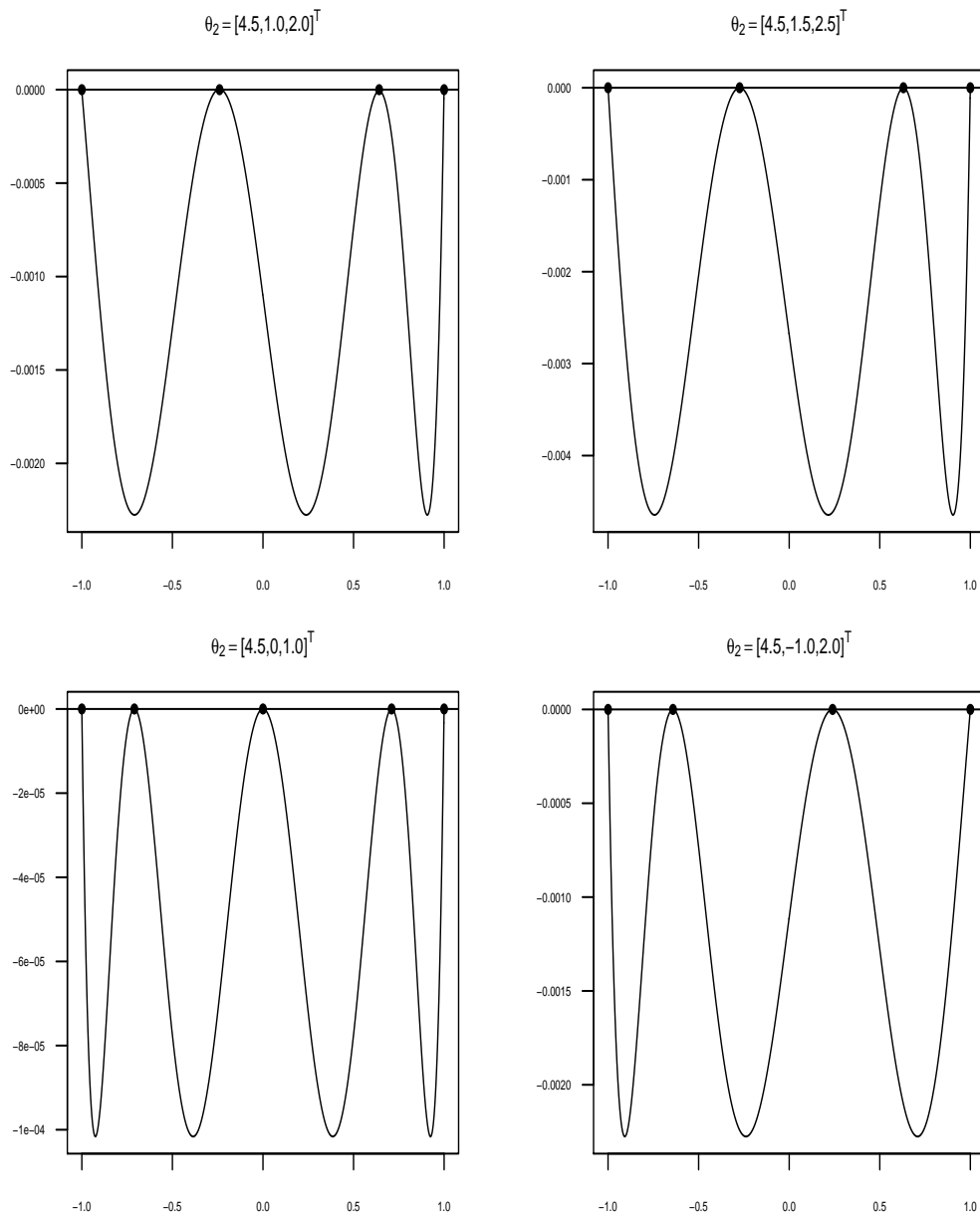


Figura 3.3: Función de Varianza para verificar  $\mathbb{T}$ -optimalidad, cuando  $\eta_2(x; \theta_2)$  es la función correcta, y  $\theta_{20} = 4.5$ .

$\theta_{21}$	$\theta_{22}$	Diseño óptimo $\xi^*$	$\tilde{\theta}_1^T$	$\Delta_2(\xi^*)$
1.0	2.0	$\begin{bmatrix} -1.000 & -0.240 & 0.640 & 1.000 \\ 0.076 & 0.272 & 0.425 & 0.227 \end{bmatrix}$	(0.91 2.26 1.36)	$2.28 \times 10^{-3}$
1.5	2.5	$\begin{bmatrix} -1.000 & -0.274 & 0.631 & 1.000 \\ 0.083 & 0.278 & 0.417 & 0.221 \end{bmatrix}$	(0.02 2.93 1.59)	$4.64 \times 10^{-3}$
-1.0	2.0	$\begin{bmatrix} -1.000 & -0.641 & 0.240 & 1.000 \\ 0.228 & 0.424 & 0.272 & 0.076 \end{bmatrix}$	(0.91 1.36 2.25)	$2.27 \times 10^{-3}$
0.0	1.0	$\begin{bmatrix} -1.000 & -0.710 & 0.000 & 0.710 & 1.000 \\ 0.044 & 0.133 & 0.258 & 0.368 & 0.198 \end{bmatrix}$	(2.67 0.92 0.92)	$1.02 \times 10^{-4}$

Tabla 3.4: Otros diseños  $\mathbb{T}$ -óptimos locales, cuando  $\eta_2(x; \theta_2)$  es la función correcta.

Modelo cierto	Distribución a priori	Diseño óptimo $\xi^*$
Modelo 1	$\pi_1$	$\begin{bmatrix} -1.00 & -0.57 & 0.43 & 1.00 \\ 0.20 & 0.36 & 0.30 & 0.14 \end{bmatrix}$
Modelo 2	$\pi_2$	$\begin{bmatrix} -1.00 & -0.32 & 0.61 & 1.00 \\ 0.09 & 0.29 & 0.41 & 0.21 \end{bmatrix}$

Tabla 3.5: Diseños  $\mathbb{T}$ -óptimos promediados por dos distribuciones a prioris

el primer modelo es el verdadero,  $\pi_1$  (ver Tabla 3.3). Además, para el caso donde el segundo modelo es el verdadero, se hallan los diseños óptimos a partir de una distribución uniforme para los cuatro valores del vector  $\theta_2$ , denotada por  $\pi_2$  (ver Tabla 3.4). Los diseños óptimos encontrados, para ambos casos, se hallan en la Tabla 3.5. Igualmente se verifica la optimalidad de los diseños a partir del respectivo teorema de equivalencia.

En este ejemplo, los diseños  $\mathbb{T}$ -óptimos obtenidos en ambos casos contienen los extremos del intervalo  $[-1, 1]$ . No existe una cota para el número de puntos de soporte asociado con este criterio, se obtuvieron diseños con cuatro puntos de soporte y se exhibió un diseño con cinco puntos de soporte. A partir del respectivo teorema de equivalencia, se verificó que los diseños obtenidos son los diseños óptimos, ya sea locales o promediados por una distribución a priori.

**Ejemplo 3.7.2.** Dos modelos para decaimiento.

A continuación se van a construir diseños  $\mathbb{T}$ -óptimos para discriminar entre los siguientes modelos para decaimiento:

$$\eta_1(x; \theta_1) = \exp(-\theta_1 x) \text{ vs. } \eta_2(x; \theta_2) = \frac{1}{1 + \theta_2 x},$$

siendo  $x \geq 0, \theta_1 \geq 0, \theta_2 \geq 0$ .



Inicialmente se supone que  $\eta_1(x; \theta_1)$  es el modelo verdadero, tomando  $\theta_1 = 1$ , es decir,  $\eta_v(x) = e^{-x}$ ,  $x \geq 0$ . Luego se tiene:

$$\Delta_2(\xi) = \min_{\theta_2} \int \left\{ e^{-x} - \frac{1}{1 + \theta_2 x} \right\}^2 d\xi(x) = \min_{\theta_2} \sum_x \left\{ e^{-x} - \frac{1}{1 + \theta_2 x} \right\}^2 \xi(x).$$

A continuación se van a hallar los diseños  $\mathbb{T}$ -óptimos para discriminar entre los dos modelos anteriores bajo el supuesto que  $0 \leq x \leq 10$ . En la Tabla 3.6 (a) se listan los diseños óptimos locales encontrados para otros valores de  $\theta_1$ , adicional a  $\theta_1 = 1.0$ .

$\theta_1$	$\theta_1 = 1$	$\theta_1 = 2$	$\theta_1 = 3$	$\theta_1 = 4$
Diseño $\xi^*$	[0.327 3.338] [0.335 0.665]	[0.163 1.669] [0.334 0.666]	[0.109 1.113] [0.334 0.666]	[0.082 0.834] [0.334 0.666]
$\tilde{\theta}_2$	1.88	3.76	5.64	7.52

(a)

$\theta_2$	$\theta_2 = 2$	$\theta_2 = 4$	$\theta_2 = 6$	$\theta_2 = 8$	$\theta_2 = 10$
Diseño $\xi^*$	[0.307 3.139] [0.334 0.666]	[0.150 1.563] [0.340 0.660]	[0.102 1.046] [0.334 0.666]	[0.077 0.785] [0.334 0.666]	[0.061 0.628] [0.449 0.551]
$\tilde{\theta}_1$	1.06	2.13	3.19	4.25	7.05

(b)

Tabla 3.6: Diseños  $\mathbb{T}$ -óptimos locales para discriminar entre los dos modelos para decaimiento, cuando: (a)  $\eta_1(x; \theta_1)$  ó (b)  $\eta_2(x; \theta_2)$  es el modelo correcto.

En el caso donde el segundo modelo es el verdadero, se encontraron los respectivos diseños  $\mathbb{T}$ -óptimos para diferentes valores de  $\theta_2$ , ver Tabla 3.6 (b). Igualmente, usando el teorema de equivalencia se verifica que en efecto los diseños obtenidos son  $\mathbb{T}$ -óptimos locales, ver la Figura 3.4 para el primer caso.

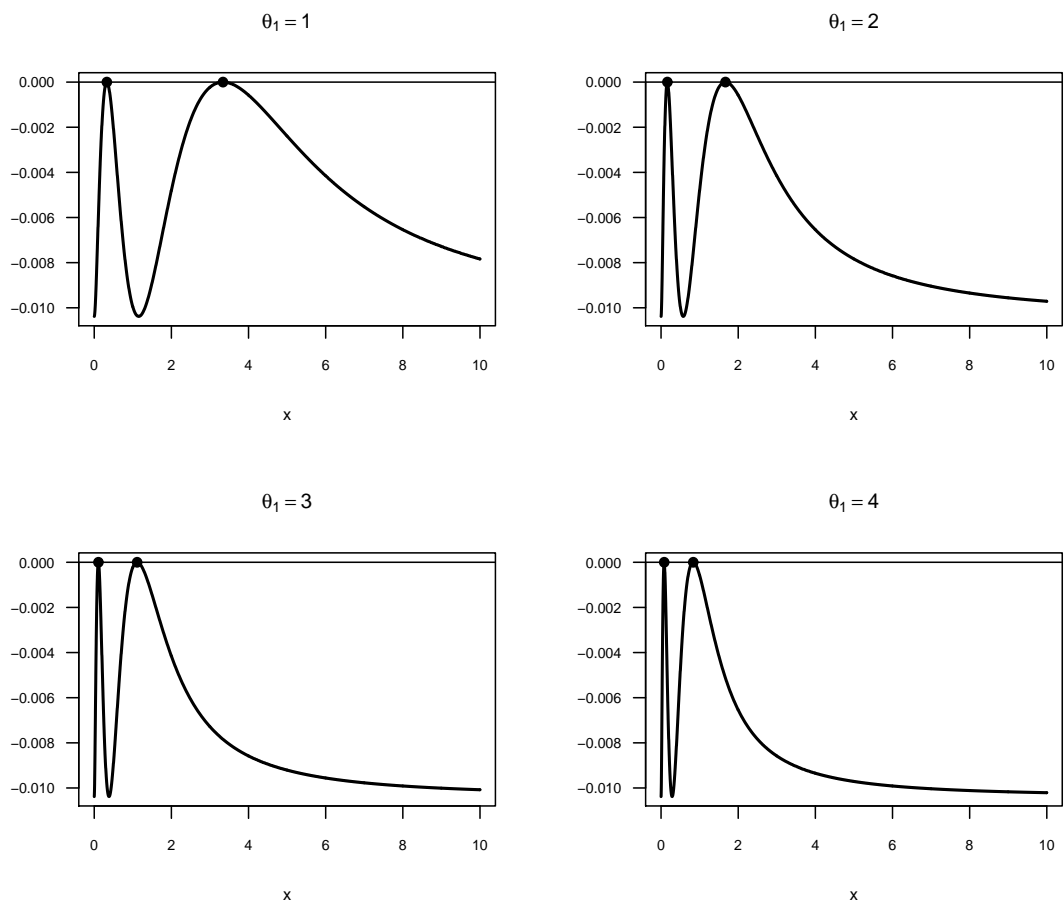
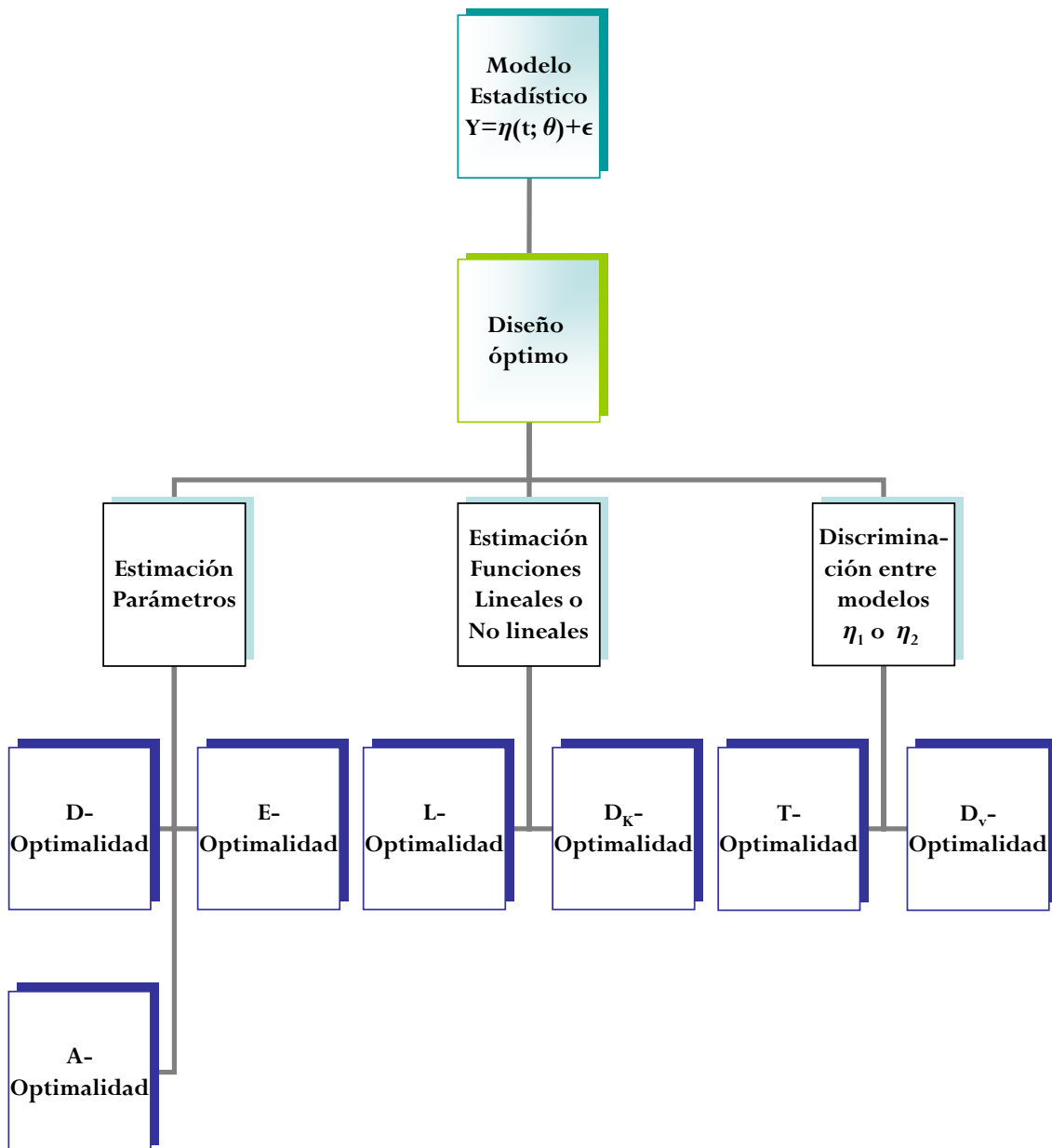


Figura 3.4: Función de Varianza para verificar  $\mathbb{T}$ -optimalidad, cuando  $\eta_1(x; \theta_1)$  es la función correcta.

### 3.8. Resumen Criterios de optimalidad

Para concluir este capítulo se presenta en el diagrama siguiente un resumen de los diferentes criterios de optimalidad organizados dependiendo si el objetivo es estimar los parámetros del modelo, o estimar funciones lineales o no lineales de los parámetros o hacer discriminación entre modelos. Todos los criterios presentados en este diagrama son de especial interés en el siguiente capítulo donde se dará la solución al problema de encontrar diseños óptimos que permitan discriminar entre dos modelos competitivos y estimar simultáneamente los parámetros y varias funciones no lineales de interés.



# Capítulo 4

## Diseños óptimos: Discriminación y estimación de funciones no lineales

### 4.1. Introducción

La búsqueda de los diseños óptimos es modelo dependiente, es decir, los diseños dependen en gran medida del modelo bajo el cual se está describiendo el comportamiento de la variable de interés. Además, como lo anotan Fedorov y Hackl (1997) en la práctica no se conoce con certeza el modelo adecuado que permita modelar el fenómeno bajo estudio. Con el fin de tener en cuenta esta dependencia en el criterio de optimalidad, se tienen varios trabajos en donde se considera una clase de modelos competitivos ya sean anidados o no y se proponen criterios de optimalidad para hallar un diseño que permita discriminar entre los modelos considerados.

Una característica no deseable que se ha observado en los diseños  $\mathbb{T}$ -óptimos para discriminar es su poca eficiencia para la estimación de los parámetros del modelo, aunado al hecho de que en varias aplicaciones el diseño resultante es singular, cómo se verá en la subsección 4.3.4. Debido a lo anterior, en varios trabajos se han propuesto criterios que tienen en cuenta tanto la discriminación entre los modelos como la estimación de los parámetros. Por ejemplo, Atkinson y Fedorov (1975b) proponen una generalización del criterio  $\mathbb{T}$ -óptimo revisado en la sección 3.7; O'Brien y Rawlings (1996) presentan una nueva propuesta para discriminar entre varios modelos competitivos no anidados y estimar simultáneamente los parámetros del modelo a partir de una combinación convexa de los logaritmos de los criterios considerados; mientras que Biswas y Chaudhuri (2002) construyen una metodología de diseños secuenciales eficientes para discriminar y estimar en modelos lineales anidados.

También, Zen y Tsai (2002) dan una solución para el problema de estimar el grado de un polinomio y simultáneamente hacer estimación de los parámetros de éste. Ellos pro-

ponen considerar la media geométrica de las eficiencias asociadas a los criterios para discriminación entre los modelos y estimación de los parámetros, respectivamente. Pukelsheim y Rosenberger (1993) presentaron un trabajo más específico para el caso polinomial. Recientemente Atkinson (2008) presenta el criterio  $\mathbb{D}\mathbb{T}$ -optimal que permite discriminar entre modelos y estimar los parámetros del modelo simultáneamente, en esencia es una combinación convexa de los criterios  $\mathbb{T}$  y  $\mathbb{D}$ -optimalidad.

En este trabajo se está interesado en proponer un criterio de optimalidad con el fin de obtener diseños óptimos que permitan discriminar entre dos posibles modelos no lineales y además, que estimen tanto el vector de parámetros como varias funciones no lineales de los parámetros,  $H_j(\theta)$ . Este tipo de problemas surge de forma natural en el área de farmacocinética, allí es de interés, además de estimar el vector de parámetros asociado con el respectivo modelo de compartimientos, estimar funciones no lineales de éste. Estas funciones sirven para entender la cinética de un medicamento bajo estudio; entre otras se consideran el área bajo la curva de concentración ( $ABC$ ), el tiempo donde se alcanza la concentración máxima ( $t_{\max}$ ) y el valor de la concentración máxima ( $\eta(t_{\max})$ ). En la sección 4.3 se ejemplifica la metodología propuesta a partir de dos modelos, uno de tres y el otro de cuatro compartimientos, con tasas de transferencia reversibles. Ambos modelos son ampliamente usados para modelar la cinética de un medicamento en un organismo. Se han realizado varios trabajos en modelos de farmacocinética usando modelos de compartimientos, ver por ejemplo Candas et al. (1988), Stroud et al. (2001), Staub et al (2003), Rajaram y Tsokos (2006). Para otros trabajos relacionados ver López-Fidalgo y Wong (2002). En esta área recientemente se han presentado, entre otros, los siguientes trabajos en diseños óptimos para modelos de compartimientos: Waterhouse et al. (2005), proponen un criterio producto para discriminar y estimar los parámetros en un modelo de compartimientos; en los trabajos de Gueorguieva, et al. (2006), Gueorguieva, et al. (2007) se presentan los diseños óptimos y un software, Popdes, para la construcción de los diseños  $\mathbb{D}$ -óptimos en modelos con respuesta multivariada para farmacocinética individual y poblacional, respectivamente.

## 4.2. Metodología Propuesta

En esta sección se propone un criterio de optimalidad a partir de los criterios usados para discriminar entre dos modelos y hacer estimación tanto del vector de parámetros como de las funciones no lineales de interés. Se consideran diferentes variantes del criterio propuesto, dependiendo del enfoque usado para discriminar entre los dos modelos considerados, y del enfoque usado para la estimación de las funciones no lineales.

Se considera el siguiente modelo de regresión no lineal general:

$$Y_i = \eta(t_i) + \epsilon_i, \quad t_i \in \chi,$$

donde las variables aleatorias  $\epsilon_i$  son independientes, normalmente distribuídas con media cero y varianza constante  $\sigma^2$ ,  $\chi \subseteq \mathbb{R}$  compacto. La función  $\eta(t)$  puede ser una de dos funciones parcialmente conocidas  $\eta_1(t; \Theta_1)$  y  $\eta_2(t; \Theta_2)$ , donde  $\Theta_1 \in \Omega_1 \subseteq \mathbb{R}^{m_1}$  y  $\Theta_2 \in \Omega_2 \subseteq \mathbb{R}^{m_2}$  son los vectores de parámetros desconocidos.

El objetivo es proponer un criterio de optimalidad que permita encontrar diseños óptimos, locales y/o promediados por una distribución a priori, con el fin de:

1. discriminar entre  $\eta_1(t; \Theta_1)$  y  $\eta_2(t; \Theta_2)$ . Y dependiendo de la decisión; que el diseño permita estimar en forma óptima:
2.  $\Theta_1$  y  $H_{1j}(\Theta_1)$ ,  $j = 1, \dots, n_l$  y/o
3.  $\Theta_2$  y  $H_{2j}(\Theta_2)$ ,  $j = 1, \dots, n_l$ ,

donde  $n_l$  es el número de funciones no lineales consideradas.

Sea  $M_i(\xi) = \int_{\chi} f_i(t; \Theta_i) f_i^T(t; \Theta_i) d\xi(t)$  la matriz de información asociada al  $i$ -ésimo modelo, con  $f_i(t; \Theta_i) = \frac{\partial \eta_i(t; \Theta_i)}{\partial \Theta_i}$ .

Defina las matrices  $K_i$  de orden  $m_i \times n_l$ , ( $i = 1, 2$ ), con la  $j$ -ésima columna de  $K_i$ , denotada por  $K_i^j$ , el gradiente de  $H_{ij}$ , es decir:

$$K_i^j = \frac{\partial H_{ij}}{\partial \Theta_i}, \quad i = 1, 2, \quad j = 1, \dots, n_l.$$

Bajo condiciones de regularidad, Seber y Wild (1989), se tiene que la matriz de varianzas-covarianzas asintótica de  $[H_{i1}(\hat{\Theta}_i) \dots H_{in_l}(\hat{\Theta}_i)]^T$  es  $K_i^T M_i^-(\xi) K_i$ , donde  $\hat{\Theta}_i$  es el estimador de mínimos cuadrados asociado al modelo  $i$ . Entonces se consideran los siguientes criterios:  $\mathbb{L}$ -optimalidad ó  $\mathbb{D}_{K_i}$ -optimalidad generalizado para el sistema  $K_i^T \Theta_i$ , ambos conducen a la minimización ya sea de la suma de las  $n_l$  varianzas asintóticas o del volumen del elipsoide asintótico de las  $n_l$ -funciones no lineales, respectivamente.

En adelante, los criterios de optimalidad cóncavos para discriminación y estimación son denotados por:

- $\phi_1(\xi)$ , para discriminación entre los dos modelos, bajo el supuesto de que el modelo correcto es:  $\eta(t; \Theta) = \eta_1(t; \Theta)$ . Al final de la demostración del teorema

de equivalencia se dá una generalización, sin necesidad de usar este supuesto considerando un criterio compuesto entre los criterios resultantes usando uno u otro modelo como correcto.

- $\phi_2(\xi) = |M_1(\xi)|^{1/m_1}$ ,  $\phi_3(\xi) = |M_2(\xi)|^{1/m_2}$ , para estimación de  $\Theta_1$  y  $\Theta_2$  respectivamente,
- $\phi_4(\xi)$ ,  $\phi_5(\xi)$ , para la estimación de las  $n_l$ -funciones no lineales en cada uno de los modelos, usando ya sea  $\eta_1$  ó  $\eta_2$ , respectivamente.

Un criterio razonable cuya finalidad sea discriminar entre los dos modelos y estimar los parámetros y funciones no lineales no debe depender de la magnitud de los criterios que entran en juego en la propuesta. Una forma de cumplir con este requerimiento es estandarizar los criterios mediante la eficiencia de los criterios individuales. Luego, usar la media geométrica de estas eficiencias como una función apropiada para combinar los distintos criterios a partir de ponderaciones adecuadas. Elegimos dar la misma ponderación a los criterios asociados a la estimación, y otra ponderación al criterio usado para la discriminación entre los dos modelos. Esto es, para  $\alpha \in [0, 1]$  se considera:

$$\psi(\xi) := \left[ \frac{\phi_1(\xi)}{\phi_1(\xi_1^*)} \right]^\alpha \left\{ \frac{\phi_2(\xi)}{\phi_2(\xi_2^*)} \frac{\phi_3(\xi)}{\phi_3(\xi_3^*)} \frac{\phi_4(\xi)}{\phi_4(\xi_4^*)} \frac{\phi_5(\xi)}{\phi_5(\xi_5^*)} \right\}^{\frac{1-\alpha}{4}},$$

donde  $\xi_j^* = \arg \max_{\xi} \phi_j(\xi)$ ,  $j = 1, \dots, 5$ .

Se pondera el criterio de discriminación por  $\alpha$ , y se supone la misma ponderación para cada una de las componentes asociadas a la estimación tanto de  $\Theta$  como de las  $n_l$  funciones no lineales,  $\frac{1-\alpha}{4}$ . Es posible tener en cuenta otras consideraciones con respecto a las ponderaciones de los criterios. La ponderación usada se considera razonable, implícitamente cuando  $\alpha > 0.2$  el criterio de discriminación tiene mayor importancia comparado con los criterios usados para las estimaciones de interés. En cualquier caso la elección del valor de  $\alpha$  dependerá de los intereses del investigador. También se puede optar por un procedimiento de selección que conduzca a hallar el  $\alpha$  que maximice el producto de las eficiencias individuales, como por ejemplo lo propone Atkinson (2008).

El criterio propuesto es un criterio de optimalidad cóncavo, luego su logaritmo también lo es, y resulta en:

$$\begin{aligned} \ln \psi(\xi) = & \alpha \log \phi_1(\xi) + \frac{1-\alpha}{4} \left\{ \frac{1}{m_1} \log |M_1(\xi)| + \frac{1}{m_2} \log |M_2(\xi)| \right. \\ & \left. + \log \phi_4(\xi) + \log \phi_5(\xi) \right\} + R, \end{aligned}$$

con  $R$  una constante que no depende del diseño  $\xi$ . Luego, en general, el criterio de optimalidad propuesto es:

$$\begin{aligned} \phi_\alpha(\xi; \Theta) = & \alpha \log \phi_1(\xi) + \frac{1-\alpha}{4} \left\{ \frac{1}{m_1} \log |M_1(\xi)| + \frac{1}{m_2} \log |M_2(\xi)| \right. \\ & \left. + \log \phi_4(\xi) + \log \phi_5(\xi) \right\}, \end{aligned} \quad (4.1)$$

Por ejemplo, si inicialmente se considera  $\mathbb{L}$ - optimalidad como criterio para estimar las funciones no lineales, es decir:  $\phi_i(\xi) = \left\{ \text{Tr} \left( K_i^T M_i^{-1}(\xi) K_i \right) \right\}^{-1}$ , entonces el criterio de optimalidad propuesto se reduce a:

$$\begin{aligned} \phi_\alpha^L(\xi; \Theta) = & \alpha \log \phi_1(\xi) + \frac{1-\alpha}{4} \left\{ \frac{1}{m_1} \log |M_1(\xi)| + \frac{1}{m_2} \log |M_2(\xi)| \right. \\ & \left. - \log \left[ \text{Tr} \left( K_1^T M_1^{-1}(\xi) K_1 \right) \right] - \log \left[ \text{Tr} \left( K_2^T M_2^{-1}(\xi) K_2 \right) \right] \right\}, \end{aligned} \quad (4.2)$$

donde  $\Theta^T = (\Theta_1^T, \Theta_2^T)$ . Es claro que este criterio depende de: el valor de  $\alpha$ , los criterios usados para discriminación y estimación de las funciones no lineales y del vector de parámetros  $\Theta$ . Se denotará este criterio por:  $\phi_\alpha^L(\xi; \Theta)$ , donde el superíndice  $L$  indica que se usó  $\mathbb{L}$ -optimalidad para la estimación de las  $n_l$  funciones no lineales en ambos modelos. Si un diseño  $\xi$  maximiza el criterio definido en 4.2, entonces se dice que  $\xi$  es  $DEFNS_\alpha^L$ -óptimo local,  $DEFNS$ , corresponde a las iniciales de: Discriminación y Estimación de Funciones No lineales Simultáneamente.

Análogamente, un diseño  $\xi$  es  $DEFNS_\alpha^D$ -óptimo local si maximiza el criterio obtenido al usar en la expresión (4.2) el criterio  $\mathbb{D}_{K_i}$ -optimalidad. Es decir,

$$\phi_i(\xi) = \left[ \det \left( K_i^T M_i^{-1}(\xi) K_i \right) \right]^{-1/n_i}.$$

El criterio resultante es:

$$\begin{aligned} \phi_\alpha^D(\xi; \Theta) = & \alpha \log \phi_1(\xi) + \frac{1-\alpha}{4} \left\{ \frac{1}{m_1} \log |M_1(\xi)| + \frac{1}{m_2} \log |M_2(\xi)| \right\} \\ & + \frac{1-\alpha}{4} \left\{ -\frac{1}{n_l} \log \left[ \det \left( K_1^T M_1^{-1}(\xi) K_1 \right) \right] - \frac{1}{n_l} \log \left[ \det \left( K_2^T M_2^{-1}(\xi) K_2 \right) \right] \right\}, \end{aligned}$$

A continuación se enuncia y demuestra el respectivo teorema de equivalencia asociado a  $\phi_\alpha^\odot(\xi; \Theta)$ , donde  $\odot$  es  $L$  ó  $D$ . Este teorema es útil para verificar si un diseño es  $DEFNS_\alpha^\odot$ -óptimo local.



**Teorema 4.2.1.** *Un diseño  $\xi$  es  $DEFNS_\alpha$ -óptimo local si y sólo sí,  $\forall t \in \chi$ ,*

$$D(t; \xi, \alpha, \Theta) = \alpha d_1(t; \xi) + \frac{1-\alpha}{4} \left\{ \frac{1}{m_1} f_1^T(t) M_1^{-1} f_1(t) + \frac{1}{m_2} f_2^T(t) M_2^{-1} f_2(t) \right. \\ \left. + d_4(t; \xi) + d_5(t; \xi) \right\} - 1 \leq 0, \quad (4.3)$$

donde  $d_1(t; \xi) = \frac{\Psi_2(t; \xi)}{\Delta_2(\xi)}$  en el caso de  $\mathbb{T}$ -óptimalidad, ver sección 3.7.

Si se usa para la estimación de las funciones no lineales:

$\mathbb{L}$ -óptimalidad entonces:

$$d_i(t; \xi) = \frac{f_j^T(t) M_j^{-1}(\xi) K_j K_j^T M_j^{-1}(\xi) f_j(t)}{\text{Tr}(K_j^T M_j^{-1}(\xi) K_j)}, \quad i = 4, 5; \quad j = i - 3,$$

y si es  $\mathbb{D}_{K_i}$ -óptimalidad:

$$d_i(t; \xi) = \frac{f_j^T(t) M_j^{-1}(\xi) K_j (K_j^T M_j^{-1}(\xi) K_j)^{-1} K_j^T M_j^{-1}(\xi) f_j(t)}{n_i}, \quad i = 4, 5; \quad j = i - 3.$$

#### Demostración:

La concavidad del criterio,  $\phi_\alpha(\xi; \Theta)$ , definido en la ecuación 4.2, se sigue del hecho que es una combinación convexa de criterios cóncavos. Luego, como los criterios involucrados son diferenciables, entonces  $\phi_\alpha$  también lo es. Ahora aplicando las propiedades de la derivada direccional vistas en la sección 3.4 se tiene:

$$D(t; \xi) = \Phi_\alpha(M(\xi); M(\delta_t)) = \alpha \Phi_1(\xi; \delta_t) + \frac{1-\alpha}{4} [\Phi_2(\xi; \delta_t) \\ + \Phi_3(\xi; \delta_t) + \Phi_4(\xi; \delta_t) + \Phi_5(\xi; \delta_t)],$$

donde:

$$\Phi_2(\xi; \delta_t) = \frac{1}{m_1} (f_1^T(t) M_1^{-1}(\xi) f_1(t) - m_1), \quad \Phi_3(\xi; \delta_t) = \frac{1}{m_2} (f_2^T(t) M_2^{-1}(\xi) f_2(t) - m_2),$$

son las derivadas direccionales de los criterios,  $\log |M_1(\xi)|^{1/m_1}$ ;  $\log |M_2(\xi)|^{1/m_2}$  en la dirección del diseño  $\delta_t$  (diseño que tiene su peso concentrado en  $t$ ). Para  $\mathbb{L}$ -óptimalidad, se tiene:

$$\Phi_4(\xi; \delta_t) = D_4(t; \xi) = \frac{f_1^T(t) M_1^{-1}(\xi) K_1 K_1^T M_1^{-1} f_1(t) - \text{Tr}(K_1^T M_1^{-1}(\xi) K_1)}{\text{Tr}(K_1^T M_1^{-1}(\xi) K_1)} = d_4(t; \xi) - 1 \\ \Phi_5(\xi; \delta_t) = D_5(t; \xi) = \frac{f_2^T(t) M_2^{-1}(\xi) K_2 K_2^T M_2^{-1} f_2(t) - \text{Tr}(K_2^T M_2^{-1}(\xi) K_2)}{\text{Tr}(K_2^T M_2^{-1}(\xi) K_2)} = d_5(t; \xi) - 1,$$

representan las derivadas direccionales en la dirección del diseño  $\delta_t$  de los criterios  $\log [\text{Tr}(K_1^T M_1^{-1}(\xi) K_1)]$  y  $\log [\text{Tr}(K_2^T M_2^{-1}(\xi) K_2)]$  respectivamente.

Para el caso de  $\mathbb{T}$ -optimalidad, al aplicar la regla de la cadena, se encuentra la siguiente derivada direccional de  $\log \Delta_2(\xi)$ :

$$\Phi_1(\xi; \delta_t) = D_1(t; \xi) = \frac{\Psi_2(t; \xi) - \Delta_2(\xi)}{\Delta_2(\xi)} = d_1(t; \xi) - 1.$$

Si se usa  $\mathbb{D}_{K_i}$ -optimalidad, es decir:  $\phi_i(\xi) = |K_{i-3}^T M_{i-3}^{-1} K_{i-3}|^{-1/n_i}$ ,  $i = 4, 5$ , se sigue del Corolario 3.5.3, página 36:

$$\Phi_4(\xi; \delta_t) = D_4(t; \xi) = \frac{f_1^T(t) M_1^{-1}(\xi) K_1 (K_1^T M_1^{-1}(\xi) K_1)^{-1} K_1^T M_1^{-1}(\xi) f_1(t) - n_l}{n_l}$$

$$\Phi_5(\xi; \delta_t) = D_5(t; \xi) = \frac{f_2^T(t) M_2^{-1}(\xi) K_2 (K_2^T M_2^{-1}(\xi) K_2)^{-1} K_2^T M_2^{-1}(\xi) f_2(t) - n_l}{n_l},$$

en ambos casos,  $d_i(t; \xi) = D_i(t; \xi) + 1$ ,  $i = 4, 5$ .

Al substituir las respectivas expresiones y después de simplificar, se obtiene la expresión del lado izquierdo de la ecuación 4.3, ya sea que se utilice  $\mathbb{L}$ - o  $\mathbb{D}_{K_i}$ -optimalidad para la estimación de las funciones no lineales.  $\square$

**Observación 4.2.1.** *El criterio propuesto depende del modelo correcto en el caso de  $\mathbb{T}$ -optimalidad. Se puede generalizar el criterio, dando ponderaciones a los criterios resultantes para discriminación, tanto cuando el primer modelo es el correcto, como cuando el segundo modelo es el correcto. Además, se puede presentar en una forma más general, suponiendo ponderaciones, no necesariamente iguales, para los criterios asociados con estimación de los parámetros y de las funciones no lineales. Es decir, el criterio se puede presentar de la siguiente forma:*

- *Considerar el producto de las eficiencias individuales y al tomar el logaritmo se obtiene:*

$$[\alpha_{11} \log \phi_{11}(\xi) + \alpha_{12} \log \phi_{12}(\xi)] + \left[ \sum_{i=2}^5 \alpha_i \log \phi_i(\xi) \right] + r,$$

donde  $r$  resulta ser una constante que no depende de  $\xi$ , y  $\alpha_{11} + \alpha_{12} + \sum_{i=2}^5 \alpha_i = 1$ . Los primeros dos sumandos están relacionados con los criterios usados para discriminar, el segundo índice indica el modelo que se asume correcto. El último sumando contiene los criterios de optimalidad para estimar los parámetros ( $\phi_3$  y  $\phi_5$ ) y las funciones no lineales en cada uno de los modelos ( $\phi_2$  y  $\phi_4$ ).

- El criterio a considerar es:

$$\phi(\xi) = [\alpha_{11} \log \phi_{11}(\xi) + \alpha_{12} \log \phi_{12}(\xi)] + \left[ \sum_{i=2}^5 \alpha_i \log \phi_i(\xi) \right].$$

- Algunas simplificaciones:

- Ponderaciones iguales para los dos criterios de discriminación:  $\beta_1 = \alpha_{11} = \alpha_{12}$ , igual ponderación para los criterios de estimación de los parámetros:  $\beta_2 = \alpha_3 = \alpha_5$ , e igual ponderación para los criterios de estimación de las funciones no lineales:  $\beta_3 = \alpha_2 = \alpha_4$ . De donde:  $\sum_{i=1}^3 2\beta_i = 1$ , luego,  $\beta_3 = \frac{1-2\beta_1-2\beta_2}{2}$ , y  $\beta_1 + \beta_2 \leq 0.5$ . El criterio a considerar es:

$$\begin{aligned} \phi_{\beta_1, \beta_2}(\xi) = & \beta_1 \log(\phi_{11}(\xi)\phi_{12}(\xi)) + \beta_2 \log(\phi_2(\xi)\phi_4(\xi)) \\ & + \frac{1-2\beta_1-2\beta_2}{2} \log(\phi_3(\xi)\phi_5(\xi)). \end{aligned}$$

- Si se supone igual ponderación para los dos criterios de discriminación e igual ponderación para los criterios de estimación de los parámetros y de las funciones no lineales, entonces se tiene:  $\beta_1 = \alpha_{11} = \alpha_{12}$ ,  $\beta_2 = \alpha_2 = \dots = \alpha_5$ , siendo  $2\beta_1 + 4\beta_2 = 1$ . Luego,  $\beta_2 = \frac{1-2\beta_1}{4}$  y  $\beta_1 \leq 0.5$ . El criterio es:

$$\phi_{\beta_1}(\xi) = \beta_1 \log[\phi_{11}(\xi)\phi_{12}(\xi)] + \frac{1-2\beta_1}{4} \log \prod_{i=2}^5 \phi_i(\xi)$$

**Observación 4.2.2.** Los modelos usados en el ejemplo, sección 4.3, como ilustración de esta propuesta son anidados, posibilitando el uso de otro criterio alternativo para discriminación,  $\mathbb{D}_s$ -optimalidad. En la sección 4.3.4 se expone este criterio más detalladamente usando el ejemplo de ilustración.

**Observación 4.2.3.** Si se conoce una distribución a priori para  $\Theta$ ,  $\pi$ , entonces el criterio general dado por la expresión 4.1 se generaliza al considerar el promedio de este criterio por alguna distribución a priori. Es decir, se considera como criterio de optimalidad con respecto a la a priori  $\pi$ :

$$\phi_{\pi}(\xi; \alpha) = \mathbb{E}_{\pi}[\phi_{\alpha}(\xi; \Theta)] = \int \phi_{\alpha}(\xi; \Theta) \pi(\Theta) d\Theta.$$

Si un diseño  $\xi$  maximiza el criterio anterior se dirá que  $\xi$  es  $DEFNS_{\alpha, \pi}$ - óptimo, usando los respectivos superíndices para indicar el criterio usado para la estimación de las funciones no lineales. En forma análoga, el respectivo teorema de equivalencia

es el resultado de promediar la función de sensibilidad dada en el Teorema 4.2.1, y se enuncia brevemente como sigue:  $\xi$  es  $DEFNS_{\alpha,\pi}$  óptimo si y solo si para cada  $t \in \chi$ :

$$\mathbb{E}_\pi [D(t; \xi, \alpha, \Theta)] \leq 0.$$

A continuación se ilustra el criterio propuesto a partir de la construcción de dos modelos estadísticos que provienen de un par de modelos de compartimientos de uso frecuente en el área de farmacocinética.

### 4.3. Ejemplo ilustrativo de la propuesta

En esta sección se ilustra el criterio de optimalidad propuesto a partir del estudio de dos modelos de compartimientos que tienen tasas de transferencia reversibles. La diferencia entre los dos modelos de compartimientos es en un compartimiento, el cual es factible que sea considerado o ignorado en el modelo estadístico bajo estudio. Ambos modelos de compartimientos son usados ampliamente en el área de farmacocinética para el estudio de la cinética de un determinado medicamento, por ejemplo en Davis et al (2007) usan el modelo de tres compartimientos para estudiar la cinética del *etodolac*, anti-inflamatorio usado para control tanto ortopédico como del dolor abdominal en caballos, cuando es administrado tanto por vía oral como intravenosa con una dosis única de  $20\text{mg/kg}$ ; mientras en Glen (2005)) referencia los resultados de dos estudios hechos con dos modelos de compartimientos, uno con dos y el otro con tres compartimientos, para el caso de administración intravenosa de dos anestésicos, *propofol* y *lidocaina*, en ambos casos se administra una dosis de  $1\text{mg/kg}$ .

En este contexto, se presenta, para cada uno de los modelos de compartimientos estudiados, la solución del respectivo sistema lineal de ecuaciones diferenciales. Después de la administración del medicamento vía oral, el compartimiento central (representando el plasma) se considera apropiado para la toma de muestras y en el modelo estadístico sólo se tiene en cuenta la concentración del medicamento en aquel compartimiento.

El objetivo se traduce en hallar diseños óptimos que discriminen entre los modelos de compartimientos considerados y que además estimen tanto las tasas de transferencia entre los diferentes compartimientos y varias funciones no lineales de interés.

Antes de cumplir con el anterior objetivo, es importante estudiar los diseños óptimos para cada uno de los modelos individualmente, y también encontrar los diseños óptimos para la discriminación entre ambos modelos. Lo anterior se hace con el fin de tener un panorama general de los diseños obtenidos individualmente. En varios de los criterios de optimalidad, además de encontrar diseños óptimos locales, también se estudian diseños óptimos promediados por diversas a priori. Las distribuciones a

prioris son definidas apropiadamente con el propósito de analizar la robustez de los diseños con respecto a la especificación de los valores locales de los parámetros. Las distribuciones a prioris consideradas se definen en la subsección 4.3.3, página 77 y en la subsección 4.3.7, página 95.

Luego se analiza el criterio propuesto en la subsección 4.3.5, para el caso de observaciones independientes donde se supone que se tiene una muestra de  $n$ -individuos que al tiempo  $t = 0$  ingerieron el medicamento, el diseño óptimo que se obtenga proporcionará los tiempos de muestreo y la respectiva proporción de individuos,  $w_i n$ , a quienes se les tomará una muestra de sangre. Sólo se extrae una muestra por individuo en un tiempo dado, representando una ventaja del procedimiento anterior, pero es válido si se tiene un gran número de individuos y la población es muy homogénea con respecto a diferentes covariables que afectan la cinética del medicamento bajo estudio. En contraste, en el Capítulo 5 se muestra un trabajo inicial con el supuesto de correlación entre las observaciones, siendo este un caso más realista, donde se tiene un solo individuo a quién se le hacen varias tomas de sangre en diferentes tiempos.

### 4.3.1. Modelo de cuatro compartimientos

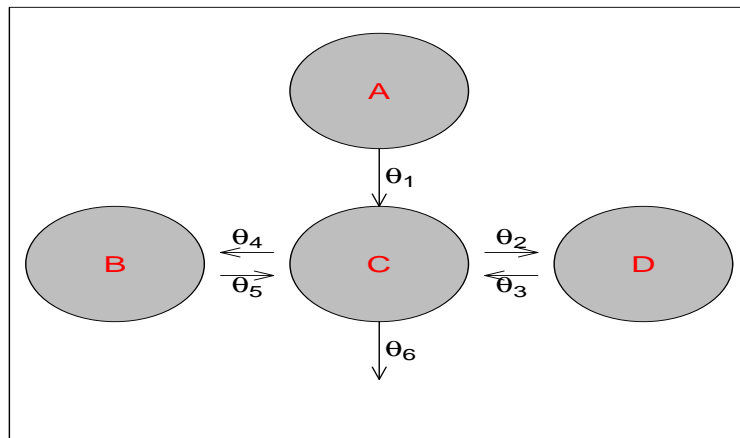


Figura 4.1: Modelo esquemático de cuatro compartimientos

A continuación, el interés va a estar orientado al estudio de uno de tales modelos, resultado de un modelo de cuatro compartimientos y con tasas de transferencia reversibles, ver Figura 4.1. El compartimiento  $C$  se considera como el compartimiento

central y los demás son llamados compartimientos periféricos. El objetivo de esta subsección es encontrar los diseños óptimos para estimar en forma óptima el vector de parámetros, formado por las tasas de transferencia entre los compartimientos. Además se hallarán diseños óptimos para la estimación de las características de interés asociadas con el compartimiento central,  $C$ . Las características de interés se listan más adelante.

En esta subsección el vector de las tasas de transferencia se denota por  $\Theta^T = (\theta_1, \theta_2, \dots, \theta_6)$ , con  $\theta_i > 0$ ,  $i = 1, \dots, 6$ . Los parámetros  $\theta_1$  y  $\theta_6$  son conocidos como las tasas de absorción y eliminación, respectivamente. Además se usa  $\eta_C(t; \Theta)$  para representar la concentración de medicamento en el compartimiento  $C$  al tiempo  $t$ .

Usando el supuesto de cinética de primer orden, el modelo de compartimientos bajo estudio tiene asociado el siguiente sistema lineal de ecuaciones diferenciales ordinarias:

$$\begin{aligned} \frac{d\eta_A(t; \Theta)}{dt} &= -\theta_1\eta_A(t; \Theta) \\ \frac{d\eta_B(t; \Theta)}{dt} &= \theta_4\eta_C(t; \Theta) - \theta_5\eta_B(t; \Theta) \\ \frac{d\eta_C(t; \Theta)}{dt} &= \theta_1\eta_A(t; \Theta) - (\theta_2 + \theta_4 + \theta_6)\eta_C(t; \Theta) + \theta_5\eta_B(t; \Theta) + \theta_3\eta_D(t; \Theta) \\ \frac{d\eta_D(t; \Theta)}{dt} &= \theta_2\eta_C(t; \Theta) - \theta_3\eta_D(t; \Theta). \end{aligned}$$

Con las siguientes condiciones iniciales:  $\eta_A(0; \Theta) = A_0$ , concentración inicial en el compartimiento de administración y  $\eta_D(0; \Theta) = \eta_B(0; \Theta) = \eta_C(0; \Theta) = 0$ , se usa el método dado en la sección 2.3.1, para encontrar explícitamente la solución del sistema anterior:

$$\begin{aligned} \eta_A(t; \Theta) &= A_0 e^{-\theta_1 t} \\ \eta_B(t; \Theta) &= A_0 \theta_1 [g_1(\Theta) e^{-\theta_1 t} + g_2(\Theta) e^{\lambda_1 t} + g_3(\Theta) e^{\lambda_2 t} + g_4(\Theta) e^{\lambda_3 t}] \\ \eta_C(t; \Theta) &= \theta_1 A_0 [h_1(\Theta) e^{-\theta_1 t} + h_2(\Theta) e^{\lambda_1 t} + h_3(\Theta) e^{\lambda_2 t} + h_4(\Theta) e^{\lambda_3 t}] \\ \eta_D(t; \Theta) &= A_0 \theta_1 [s_1(\Theta) e^{-\theta_1 t} + s_2(\Theta) e^{\lambda_1 t} + s_3(\Theta) e^{\lambda_2 t} + s_4(\Theta) e^{\lambda_3 t}], \end{aligned} \quad (4.4)$$

donde:

1.  $\lambda_i$ ,  $i = 1, 2, 3$  son las soluciones de la ecuación cúbica:

$$\lambda^3 + \lambda^2 (\theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6) + \lambda ((\theta_2 + \theta_3 + \theta_6) \theta_5 + (\theta_4 + \theta_6) \theta_3) + \theta_3 \theta_5 \theta_6 = 0,$$

por tanto las  $\lambda_i$ ,  $i = 1, 2, 3$ , satisfacen las siguientes relaciones:

- a)  $-\sum_{i=1}^3 \lambda_i = \sum_{i=2}^6 \theta_i$   
 b)  $\lambda_3(\lambda_1 + \lambda_2) + \lambda_1\lambda_2 = (\theta_2 + \theta_3 + \theta_6)\theta_5 + \theta_3(\theta_4 + \theta_6)$   
 c)  $-\lambda_1\lambda_2\lambda_3 = \theta_3\theta_5\theta_6$ .

Para valores razonables de las tasas de transferencia,  $\theta_i > 0$ , se ha mostrado con simulación, que las tres soluciones anteriores son reales y además negativas.

2. Las funciones  $h_i(\Theta)$ ,  $i = 1, \dots, 4$ , están dadas por las siguientes expresiones:

$$\begin{aligned} h_2(\Theta) &= \frac{(\lambda_1 + \theta_3)(\lambda_1 + \theta_5)}{(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)(\lambda_1 + \theta_1)} \\ h_3(\Theta) &= -\frac{(\lambda_2 + \theta_3)(\lambda_2 + \theta_5)}{(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_2 + \theta_1)} \\ h_4(\Theta) &= \frac{(\lambda_3 + \theta_3)(\lambda_3 + \theta_5)}{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3)(\lambda_3 + \theta_1)} \\ h_1(\Theta) &= -h_2(\Theta) - h_3(\Theta) - h_4(\Theta). \end{aligned}$$

Las otras funciones,  $g_i(\Theta)$ ,  $s_i(\Theta)$ ,  $i = 1, 2, 3, 4$ , son funciones de  $\Theta$ , en lo que sigue no se usan y por brevedad las expresiones explícitas de estas funciones no se consideran.

El modelo estadístico, que describe la concentración en el compartimiento  $C$ , está formado por una componente determinística,  $\eta_C(t; \Theta)$  y una componente aleatoria, es decir:

$$Y(t_i; \Theta) = \eta_C(t_i; \Theta) + \epsilon_i, \quad i = 1, \dots, N, \quad t_i \in [0, t_\infty], \quad (4.5)$$

donde  $\epsilon_i$  son variables aleatorias normales con media cero y varianza constante  $\sigma^2$ , y la expresión para  $\eta_C(t_i; \Theta)$  está dada en la ecuación (4.4). El símbolo  $t_\infty$  denota el tiempo para el cual es razonable suponer que la concentración es cercana a cero. En adelante el modelo dado por la ecuación 4.5 se llamará modelo  $I$ , y el intervalo  $[0, t_\infty]$  se denotará por  $\chi$

La matriz de información asociada con el modelo  $I$ , ecuación 4.5, es:

$$M(\xi; \Theta) = \int_{\chi} f(t; \Theta) f^T(t; \Theta) d\xi(t),$$

donde  $\chi$ , y  $f(t; \Theta)$  es el gradiente de la función  $\eta_C(t; \Theta)$ , y  $\xi$  es un diseño con  $s$ -puntos de soporte, representado por:

$$\xi = \begin{bmatrix} t_1 & \dots & t_s \\ w_1 & \dots & w_s \end{bmatrix},$$

siendo  $w_i = \xi(t_i)$ , donde  $\sum_i w_i = 1$ , y  $t_i \in \mathcal{X}$ .

Para estimar en forma óptima a  $\Theta$ , se usa el criterio del determinante. Debido a que la matriz de información depende de  $\Theta$ , se hallan diseños óptimos locales para un valor a priori de  $\Theta$ ,  $\Theta_0$ . El valor de  $\Theta_0$  lo elige el experimentador a partir de su experiencia, estimación de experimentos previos. Se busca el diseño  $\xi$  que maximiza  $|M(\xi; \Theta_0)|^{1/6}$  o equivalentemente, aquel diseño que maximiza  $\log |M(\xi, \Theta_0)|$ . En la siguiente sección se construyen estos diseños.

### Diseños $\mathbb{D}$ -óptimos locales

Como ilustración se tomó el siguiente valor local para  $\Theta$ :

$$\Theta_0^T = (0.30, 0.20, 0.15, 0.05, 0.08, 0.25).$$

seleccionado al azar entre diferentes posibilidades. Además, se consideró el valor de  $A_0$  fijo,  $A_0 = 4$ .

En la Figura 4.2 se muestra la curva para la concentración en el compartimiento  $C$ , con el valor  $\Theta_0$ ,  $A_0 = 4$ .

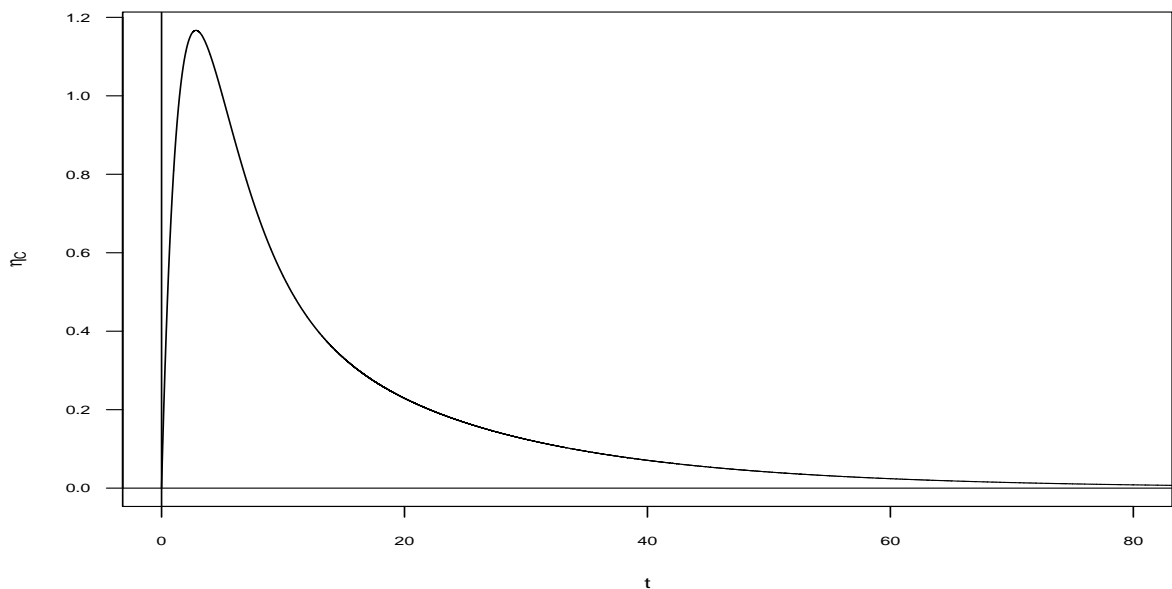


Figura 4.2: Concentración en el compartimiento  $C$  para el valor local de  $\Theta$ ,  $\Theta_0$ .



$\Theta$	Diseño óptimo ( $\xi$ )	Valor óptimo ( $ M(\xi) ^{1/6}$ )												
$\Theta_0$	<table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>1.0865</td> <td>3.8216</td> <td>9.0540</td> <td>18.5385</td> <td>35.5952</td> <td>67.8752</td> </tr> <tr> <td>0.1667</td> <td>0.1667</td> <td>0.1667</td> <td>0.1667</td> <td>0.1667</td> <td>0.1667</td> </tr> </table>	1.0865	3.8216	9.0540	18.5385	35.5952	67.8752	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	0.0316
1.0865	3.8216	9.0540	18.5385	35.5952	67.8752									
0.1667	0.1667	0.1667	0.1667	0.1667	0.1667									

 Tabla 4.1: Diseño  $\mathbb{D}$ -óptimo local, modelo I.

A partir de este gráfico se ve razonable considerar  $t_\infty = 100$ .

El diseño óptimo se halla numéricamente implementando el algoritmo dado en el Apéndice A, adaptado para  $\mathbb{D}$ -optimalidad. El diseño resultante se muestra en la Tabla 4.1. El diseño permite estimar  $\Theta$  en forma óptima, si se toman observaciones en los siguientes tiempos: 1.08, 3.82, 9.05, 18.54, 35.60 y en 67.88.

Con el teorema de equivalencia de Kiefer y Wolfowitz (1960), gráficamente se verifica que en efecto el diseño obtenido es  $\mathbb{D}$ -óptimo local, ver la Figura 4.3. Se observa que para cualquier  $t \in [0, 100]$ :

$$f^T(t; \Theta_0)M^{-1}(\xi; \Theta_0)f(t; \Theta_0) - 6 \leq 0,$$

y además en los puntos de soporte del diseño óptimo la función de sensibilidad o varianza es cero.

En la siguiente sección se encuentran los diseños óptimos para la estimación de las funciones no lineales de interés.

### Diseños $\mathbb{L}$ -óptimos locales

En esta sección se muestran los diseños  $\mathbb{L}$ -óptimos locales obtenidos para estimar diversas características de interés, tales como: el área bajo la curva de concentración, el tiempo para la concentración máxima, la concentración máxima. Además de las cantidades de interés mencionadas antes se consideró relevante estimar el primer tiempo,  $T_{0.5}$ , donde se alcanza un 50% de la concentración máxima, es decir,

$$\eta_C(T_{0.5}; \Theta_0) = 0.50\eta_C(t_{\text{máx}}; \Theta_0).$$

Cada una de las cantidades anteriores son funciones no lineales de  $\Theta$ . También, en esta sección, se consideró pertinente la estimación de dos cantidades lineales: La diferencia entre las tasas reversibles:  $\theta_4 - \theta_5$  y la tasa de absorción  $\theta_1$ . Inicialmente se halla un diseño  $\mathbb{L}$ -óptimo local, tomando como valor a priori el valor de  $\Theta_0$  dado en la subsección previa. Es decir, se necesita un diseño  $\xi$  que minimice:

$$\text{Tr} \{K^T(\Theta_0)M^{-1}(\xi; \Theta_0)K(\Theta_0)\},$$

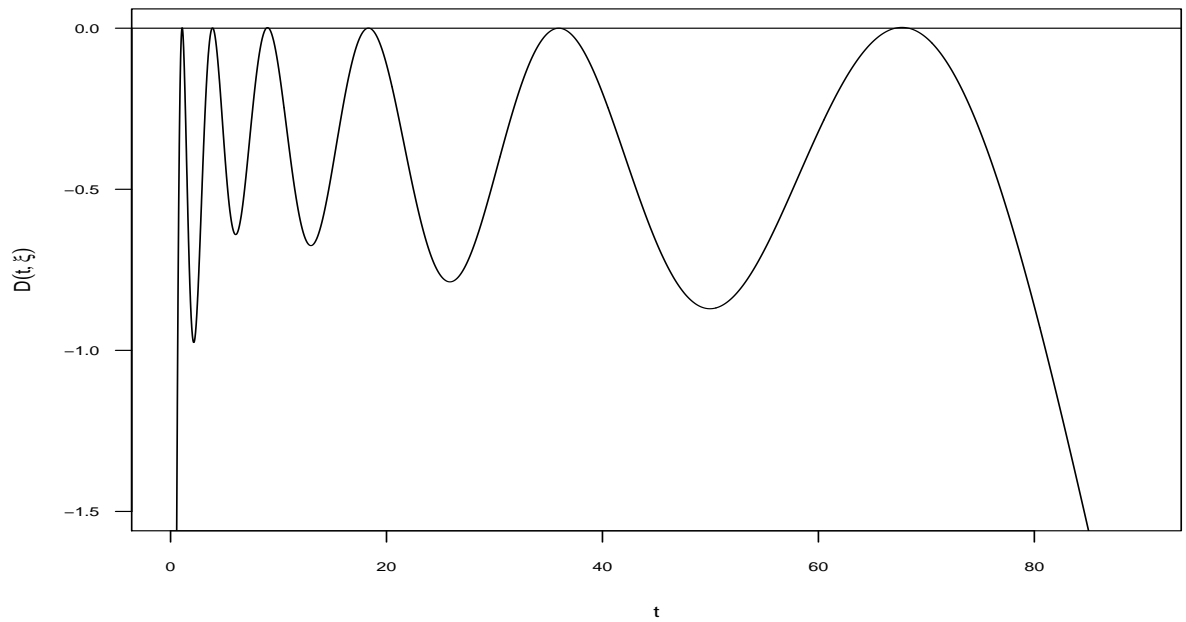


Figura 4.3: Función de varianza para  $\mathbb{D}$ -optimalidad local, modelo  $I$ .

o equivalentemente, maximice:

$$\left\{ \text{Tr} \left[ K^T(\Theta_0) M^{-1}(\xi; \Theta_0) K(\Theta_0) \right] \right\}^{-1},$$

o en forma equivalente si maximiza:

$$-\log \left\{ \text{Tr} \left[ K^T(\Theta_0) M^{-1}(\xi; \Theta_0) K(\Theta_0) \right] \right\},$$

donde las columnas de  $K$  contienen los gradientes de las cantidades que se desean estimar:

$$K = [\nabla H_1(\Theta_0) \quad \nabla H_2(\Theta_0) \quad \nabla H_3(\Theta_0) \quad \nabla H_4(\Theta_0) \quad \nabla H_5(\Theta_0) \quad \nabla H_6(\Theta_0)],$$

donde:

- $H_1(\Theta_0) = \text{Area} = \int_0^\infty \eta_C(t; \Theta_0) dt = \frac{A_0}{\theta_6}$ .
- $H_2(\Theta_0) = t_{\text{máx}}$ , tiempo para la concentración máxima.
- $H_3(\Theta_0) = \eta_C(t_{\text{máx}}; \Theta_0)$ , concentración máxima.
- $H_4(\Theta_0) = T_{0.50}$ , primer tiempo para alcanzar un 50% de la concentración máxima.
- $H_5(\Theta_0) = \theta_4 - \theta_5$ .
- $H_6(\Theta_0) = \theta_1$ .

En el Apéndice B.1 se dan las expresiones explícitas de los gradientes para cada una de las funciones anteriores.

Numéricamente se encuentra el diseño  $\mathbb{L}$ -óptimo local para estimar las características de interés. El diseño obtenido da como tiempos de muestreo aproximados: 0.74, 3.57, 8.99, 19.03, 42.31, 72.13, en las unidades apropiadas ( horas, minutos ), con frecuencias respectivas: 0.0467, 0.0548, 0.07604, 0.1677, 0.2261, 0.4287. Por ejemplo, para una muestra de tamaño  $n$ , aproximadamente el 5% de las observaciones se toma en el primer tiempo, 0.73, aumentando el porcentaje a medida que se consideran tiempos de muestreo más grandes. En la sección 5.2 se muestra un procedimiento de redondeo eficiente para diseños aproximados, da el número de observaciones para cada tiempo óptimo de muestreo.

Se verifica gráficamente que el diseño obtenido,  $\xi$ , es  $\mathbb{L}$ -óptimo local, observe que para cualquier  $t \in [0, 100]$  se cumple que:

$$f^T(t; \Theta_0) M^{-1}(\xi; \Theta_0) K(\Theta_0) K^T(\Theta_0) M^{-1}(\xi; \Theta_0) f(t; \Theta_0) - \text{Tr} \left( K^T(\Theta_0) M^{-1}(\xi; \Theta_0) K(\Theta_0) \right) \leq 0,$$

además, la función anterior se hace cero en los puntos de soporte del diseño óptimo.

$\Theta$	Diseño óptimo ( $\xi$ )	$\{\text{Tr}(K^T M^{-1}(\xi)K)\}^{-1}$												
$\Theta_0$	<table border="1"> <tr> <td>0.7395</td> <td>3.5663</td> <td>8.9894</td> <td>19.0320</td> <td>42.3118</td> <td>72.1318</td> </tr> <tr> <td>0.0467</td> <td>0.0548</td> <td>0.0760</td> <td>0.1677</td> <td>0.2261</td> <td>0.4287</td> </tr> </table>	0.7395	3.5663	8.9894	19.0320	42.3118	72.1318	0.0467	0.0548	0.0760	0.1677	0.2261	0.4287	$4.766 \times 10^{-5}$
0.7395	3.5663	8.9894	19.0320	42.3118	72.1318									
0.0467	0.0548	0.0760	0.1677	0.2261	0.4287									

Tabla 4.2: Diseños  $\mathbb{L}$ -óptimos locales, modelo  $I$ .

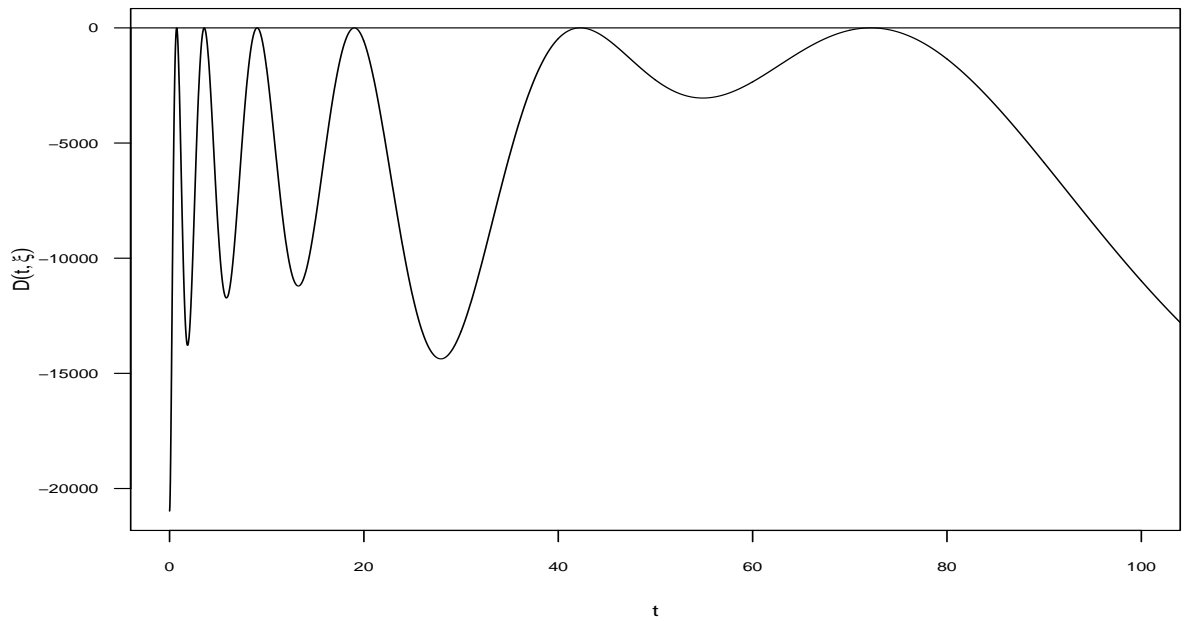


Figura 4.4: Función de varianza para  $\mathbb{L}$ -optimalidad local, modelo  $I$ .

### 4.3.2. Modelo de tres compartimientos

En esta sección se estudia un modelo de tres compartimientos como alternativa al modelo estudiado en la sección anterior, difiere por la no presencia de uno de los compartimientos periféricos, ver Figura 4.5. A continuación, igual al modelo de cuatro compartimientos, se encuentra la expresión de la concentración del medicamento en el compartimiento central. Luego se hallan los diseños  $\mathbb{D}$  y  $\mathbb{L}$ -óptimos locales para estimar  $\mathbb{B}$ , vector de tasas de transferencia asociada al modelo bajo estudio, y estimación de funciones no lineales, respectivamente.

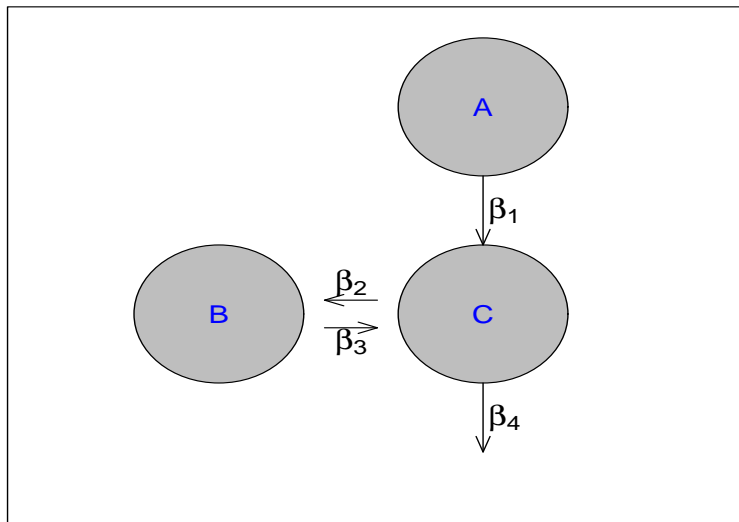


Figura 4.5: Modelo esquemático para tres compartimientos

En esta sección, el vector de las tasas de transferencia es denotado por:  $\mathbb{B}^T = (\beta_1, \beta_2, \beta_3, \beta_4)$ . En forma análoga a lo hecho para el modelo anterior, con cinética de primer orden, se halla el siguiente sistema de ecuaciones diferenciales ordinarias

asociado con la Figura 4.5:

$$\begin{aligned}\frac{d\eta_A(t; \mathbb{B})}{dt} &= -\beta_1\eta_A(t; \mathbb{B}) \\ \frac{d\eta_C(t; \mathbb{B})}{dt} &= \beta_1\eta_A(t; \mathbb{B}) + \beta_3\eta_B(t; \mathbb{B}) - (\beta_4 + \beta_2)\eta_C(t; \mathbb{B}) \\ \frac{d\eta_B(t; \mathbb{B})}{dt} &= \beta_2\eta_C(t; \mathbb{B}) - \beta_3\eta_B(t; \mathbb{B}).\end{aligned}$$

Con las condiciones iniciales:  $\eta_A(0; \mathbb{B}) = A_0$  y  $\eta_B(0; \mathbb{B}) = \eta_C(0; \mathbb{B}) = 0$ , en la sección 2.3.1 se mostró la solución del sistema anterior dada por:

$$\begin{aligned}\eta_A(t; \mathbb{B}) &= A_0 e^{-\beta_1 t} \\ \eta_B(t; \mathbb{B}) &= q_1(\mathbb{B}) e^{-\beta_1 t} + q_2(\mathbb{B}) e^{\lambda_1 t} + q_3(\mathbb{B}) e^{\lambda_2 t} \\ \eta_C(t; \mathbb{B}) &= \frac{A_0 \beta_1}{\lambda_1 - \lambda_2} [r_1(\mathbb{B}) e^{-\beta_1 t} + r_2(\mathbb{B}) e^{\lambda_1 t} + r_3(\mathbb{B}) e^{\lambda_2 t}],\end{aligned}$$

donde:

1. las funciones  $\lambda_i = \lambda_i(\mathbb{B})$ ,  $i = 1, 2$ , son soluciones de la ecuación cuadrática:

$$\lambda^2 + \lambda(\beta_2 + \beta_3 + \beta_4) + \beta_3\beta_4 = 0,$$

y por tanto satisfacen las siguientes relaciones:

- a)  $\lambda_1 \lambda_2 = \beta_3 \beta_4$ .
- b)  $\lambda_1 + \lambda_2 = -(\beta_2 + \beta_3 + \beta_4)$ .
- c) Se cumple que  $\lambda_1$  y  $\lambda_2$  son raíces negativas y distintas, para  $\beta_i > 0$  y diferentes.

2. Las funciones  $r_i(\mathbb{B})$ ,  $i = 1, 2, 3$  están dadas por:

$$\begin{aligned}r_2(\mathbb{B}) &= \frac{\lambda_1 + \beta_3}{\beta_1 + \lambda_1} \\ r_3(\mathbb{B}) &= -\frac{\lambda_2 + \beta_3}{\beta_1 + \lambda_2} \\ r_1(\mathbb{B}) &= -(r_2(\mathbb{B}) + r_3(\mathbb{B})).\end{aligned}$$

Las funciones  $q_i(\mathbb{B})$ ,  $i = 1, 2, 3$ , son funciones no lineales de  $\mathbb{B}$ . Por brevedad no se dan las expresiones explícitas.

De manera análoga, el modelo estadístico, que describe la concentración en el compartimiento  $C$ , es:

$$\begin{aligned} Y(t_i; \mathbb{B}) &= \eta_C(t_i; \mathbb{B}) + \epsilon_i, \quad i = 1, \dots, N. \\ &= \frac{A_0 \beta_1}{\lambda_1 - \lambda_2} [r_1(\mathbb{B}) e^{-\beta_1 t} + r_2(\mathbb{B}) e^{\lambda_1 t} + r_3(\mathbb{B}) e^{\lambda_2 t}] + \epsilon, \end{aligned} \quad (4.6)$$

donde las  $\epsilon_i$  cumplen los mismos supuestos de 4.5. En adelante, este modelo se llamará modelo  $II$ .

La matriz de información asociada con el modelo 4.6, se expresa como:

$$M(\xi; \mathbb{B}) = \int_{\chi} w(t; \mathbb{B}) w^T(t; \mathbb{B}) d\xi(t),$$

donde  $w(t; \mathbb{B})$  es el gradiente de  $\eta_C(t; \mathbb{B})$ , ver Apéndice B.3, y  $\chi = [0, t_\infty]$ .

### Diseños $\mathbb{D}$ -óptimos locales

Se usa el criterio del determinante para estimar las tasas de transferencia asociadas con el modelo 4.6. En este caso, un diseño  $\xi$  es  $\mathbb{D}$ -óptimo si y solo si maximiza  $|M(\xi; \mathbb{B}_0)|^{1/4}$  equivalentemente si maximiza:

$$\log |M(\xi; \mathbb{B}_0)|,$$

donde  $\mathbb{B}_0$  es un valor a priori de  $\mathbb{B}$ .

Con propósitos ilustrativos se tomó el siguiente valor local de  $\mathbb{B}$ ,

$$\mathbb{B}_0^T = (0.40, 0.28, 0.10, 0.30),$$

y se dejó el valor de  $A_0$  fijo en 4.0.

En la figura 4.6 se muestra la curva de concentración para el compartimiento  $C$ .

En la Tabla 4.3 se exhibe el diseño  $\mathbb{D}$ -óptimo local para la estimación del vector  $\mathbb{B}$ . Con este diseño los tiempos de muestreo son 1.15, 4.14, 11.07 y 33.70, con la misma frecuencia de observaciones en cada uno de ellos.

En la Figura 4.7 se verifica la  $\mathbb{D}$ -optimalidad del diseño obtenido, es decir, para cualquier  $t \in [0, 100]$  se tiene que:  $w^T(t; \mathbb{B}_0) M^{-1}(\xi; \mathbb{B}_0) w(t; \mathbb{B}_0) - 4 \leq 0$ .

### Diseños $\mathbb{L}$ -óptimos locales

Con el fin de ilustrar la construcción de los diseños  $\mathbb{L}$ -óptimos locales para el modelo en consideración, se consideró el mismo valor local usado para los diseños  $\mathbb{D}$ -óptimos. El objetivo es encontrar diseños  $\mathbb{L}$ -óptimos locales que permitan estimar en forma óptima:

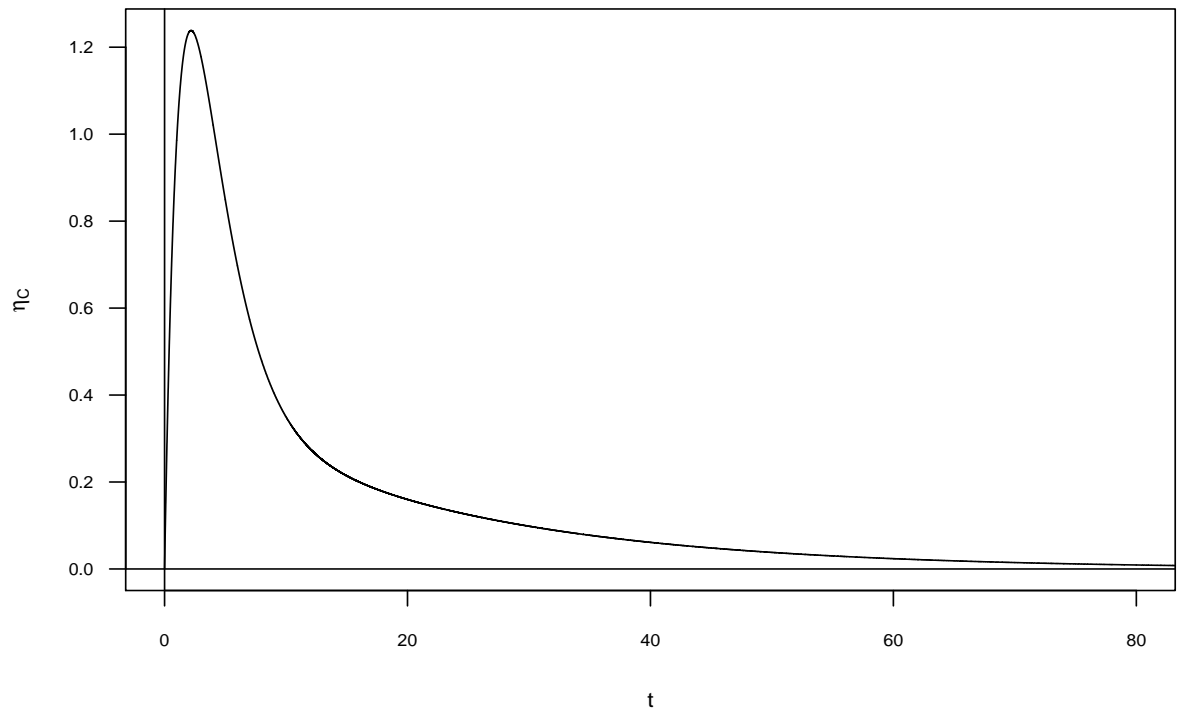


Figura 4.6: Concentración en el compartimiento  $C$  para el valor local  $\mathbb{B}_0$ ,  $A_0 = 4.0$ .

$\mathbb{B}$	Diseño óptimo ( $\xi$ )	Valor óptimo $ M(\xi) ^{1/4}$								
$\mathbb{B}_0$	<table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>1.1443</td> <td>4.1087</td> <td>11.0067</td> <td>33.6269</td> </tr> <tr> <td>0.2500</td> <td>0.2500</td> <td>0.2500</td> <td>0.2500</td> </tr> </table>	1.1443	4.1087	11.0067	33.6269	0.2500	0.2500	0.2500	0.2500	0.402
1.1443	4.1087	11.0067	33.6269							
0.2500	0.2500	0.2500	0.2500							

Tabla 4.3: Diseño  $\mathbb{D}$ -óptimo local, modelo  $II$ .



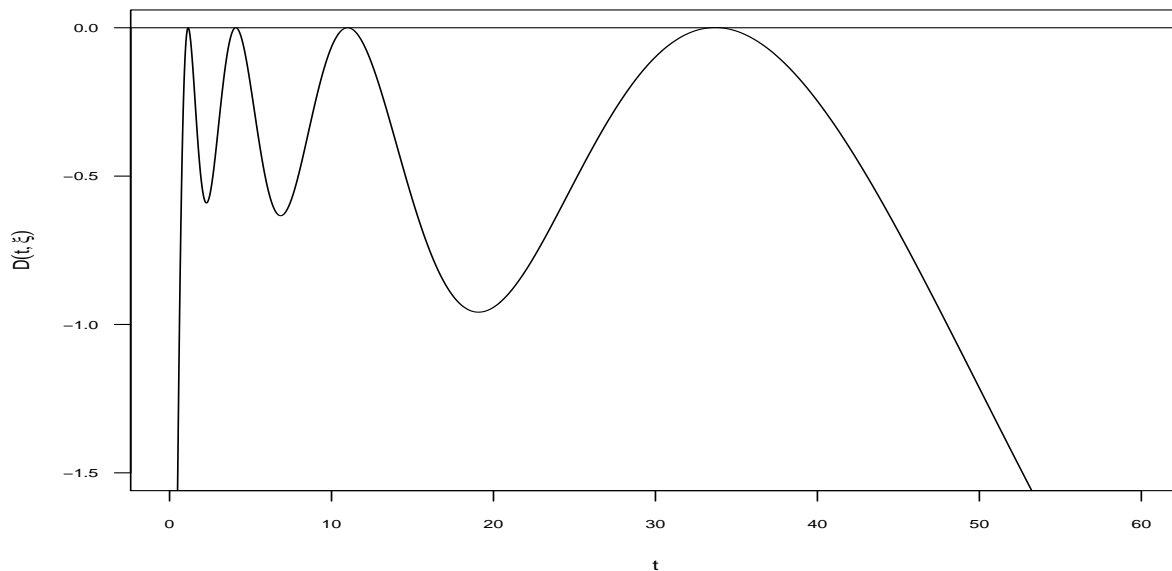


Figura 4.7: Función de varianza para  $\mathbb{D}$ -optimalidad local, modelo *II*.

1.  $G_1(\mathbb{B}) = \int_0^\infty \eta_C(t; \mathbb{B}) dt = \frac{A_0}{\beta_4}$ , área bajo la curva de concentración,
2.  $G_2(\mathbb{B}) = t_{\text{máx}}$ , tiempo para la concentración máxima,
3.  $G_3(\mathbb{B}) = \eta_C(t_{\text{máx}}; \mathbb{B})$ , concentración máxima,
4.  $G_4(\mathbb{B}) = T_{0.5}$ , primer tiempo donde la concentración alcanza el 50% de la concentración máxima.

En el Apéndice B.3 se encuentran los respectivos gradientes para las funciones no lineales,  $G_i(\mathbb{B})$ . Se eligió  $\mathbb{L}$ -optimalidad como un criterio apropiado para estimar en forma conjunta las cantidades anteriores, también se puede usar  $\mathbb{D}_K$ -optimalidad, pero en este caso coinciden ambos diseños usando los criterios  $\mathbb{D}$  y  $\mathbb{L}$ -optimalidad. El diseño  $\mathbb{L}$ -óptimo local para  $\mathbb{B}_0$  se halla tal que minimiza:

$$\text{Tr} [K^T(\mathbb{B}_0)M^{-1}(\xi; \mathbb{B}_0)K(\mathbb{B}_0)],$$

o equivalentemente maximiza:

$$\{\text{Tr} [K^T(\mathbb{B}_0)M^{-1}(\xi; \mathbb{B}_0)K(\mathbb{B}_0)]\}^{-1},$$

donde las columnas de la matriz  $K$  son los gradientes de las cantidades que se desean estimar:

$$K = [\nabla G_1(\mathbb{B}_0) \quad \nabla G_2(\mathbb{B}_0) \quad \nabla G_3(\mathbb{B}_0) \quad \nabla G_4(\mathbb{B}_0)].$$

$\mathbb{B}$	Diseño óptimo ( $\xi$ )	Valor óptimo $\{Tr(K^T M^{-1}(\xi)K)\}^{-1}$
$\mathbb{B}_0$	$\begin{bmatrix} 0.9848 & 4.2209 & 13.1466 & 30.2916 \\ 0.0286 & 0.1079 & 0.0087 & 0.8548 \end{bmatrix}$	4597.238

Tabla 4.4: Diseño  $\mathbb{L}$ -óptimo local, modelo *II*.

Numéricamente se encuentra el diseño  $\mathbb{L}$ -óptimo local, ver Tabla 4.4. En la Figura 4.8 se verifica que en efecto el diseño obtenido es  $\mathbb{L}$ -óptimo local, es decir, se muestra que para cualquier  $t \in [0, 100]$  se cumple que:

$$w^T(t; \mathbb{B}_0)M^{-1}(\xi; \mathbb{B}_0)K(\mathbb{B}_0)K^T(\mathbb{B}_0)M^{-1}(\xi; \mathbb{B}_0)w(t; \mathbb{B}_0) - Tr(K^T(\mathbb{B}_0)M^{-1}(\xi; \mathbb{B}_0)K(\mathbb{B}_0)) \leq 0,$$

y además, la función anterior es cero en los puntos de soporte del diseño óptimo.

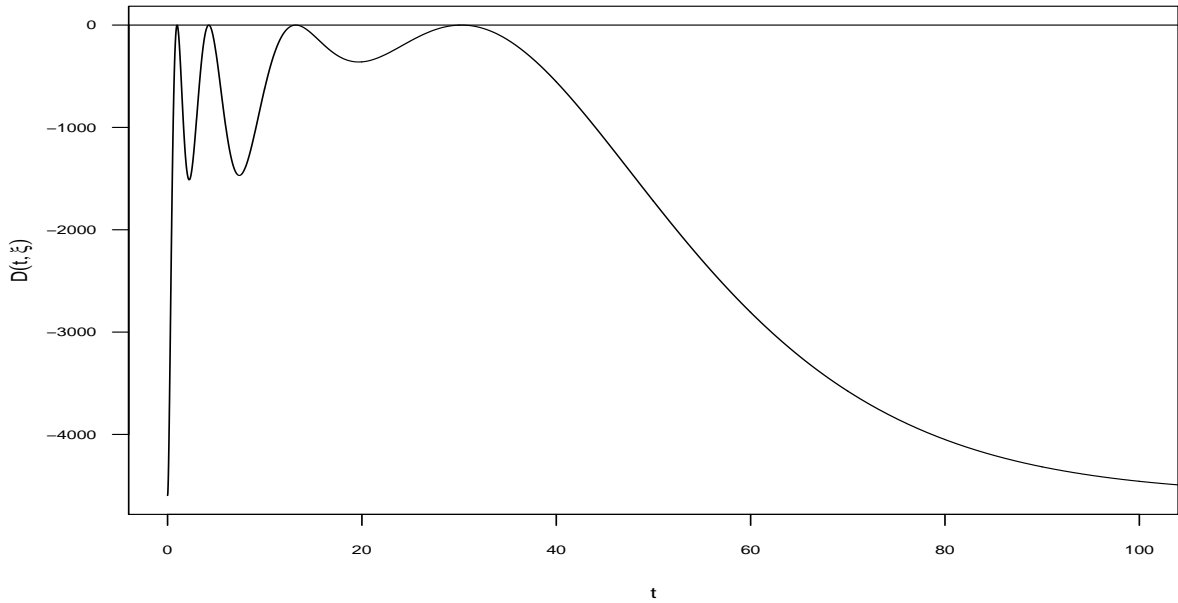


Figura 4.8: Función de varianza para  $\mathbb{L}$ -optimalidad local, modelo *II*.

En la siguiente sección se definen diferentes distribuciones a priori que varían alrededor del valor local. Se usan con el fin de analizar la robustez de los diseños óptimos frente a variaciones en los parámetros locales individuales.

### 4.3.3. Diseños óptimos promediados por una distribución a priori

Esta clase de diseños son una alternativa de solución para el problema de especificación del valor local para  $\Theta \in \mathbb{R}^m$ . En este caso se necesita especificar una distribución a priori para el vector de parámetros, denotada por  $\pi(\Theta)$ , y se considera como criterio de optimalidad, el promedio del respectivo criterio de optimalidad con respecto a la distribución a priori considerada. Luego se halla el diseño que minimice dicho criterio. Se hizo un primer trabajo con las distribuciones a prioris definidas en la subsección 4.3.3 ver López y Ramos (2007b) usando otro modelo de compartimientos.

Para el modelo  $I$ , se analizan los siguientes criterios:

1. El diseño  $\xi^*$  es un diseño  $\mathbb{D}$ -óptimo promediado por la distribución a priori  $\pi$ , abreviado por  $\mathbb{D}_\pi$ -óptimo si:

$$\begin{aligned} \xi^* &= \arg \min_{\xi} \mathbb{E}_{\pi} [-\log |M(\xi; \Theta)|] \\ &= \arg \min_{\xi} \int -\log |M(\xi; \Theta)| \pi(\Theta) d\Theta \end{aligned}$$

De nuevo, al aplicar el teorema de equivalencia, Fedorov y Hackl (1997), se tiene:  $\xi^*$  es un diseño  $\mathbb{D}_\pi$ -óptimo si y sólo si:

$$\forall t \in [0, t_{\infty}] \quad \mathbb{E}_{\pi} [f^T(t; \Theta) M^{-1}(\xi; \Theta) f(t; \Theta)] - m \leq 0;$$

donde  $m$  es el número de componentes del vector  $\Theta$ .

2. El diseño  $\xi^*$  es un diseño  $\mathbb{L}$ -óptimo promediado por la distribución a priori  $\pi$ , abreviado por  $\mathbb{L}_\pi$ -óptimo, si:

$$\begin{aligned} \xi^* &= \arg \min_{\xi} \mathbb{E}_{\pi} [\text{Tr} (K^T(\Theta) M^{-1}(\xi; \Theta) K(\Theta))] \\ &= \arg \min_{\xi} \int \text{Tr} (K^T(\Theta) M^{-1}(\xi; \Theta) K(\Theta)) \pi(\Theta) d\Theta \end{aligned}$$

Al aplicar el respectivo teorema de equivalencia, Fedorov y Hackl (1997), se tiene que:  $\xi^*$  es un diseño  $\mathbb{L}_\pi$ -óptimo si y sólo si:

$$\forall t \in [0, t_{\infty}] \quad \mathbb{E}_{\pi} [f^T(t) M^{-1}(\xi^*) K K^T M^{-1}(\xi^*) f(t) - \text{Tr} (K^T M^{-1}(\xi^*) K)] \leq 0.$$

3. El diseño  $\xi^*$  es un diseño  $\log \mathbb{L}$ -óptimo promediado por la distribución a priori  $\pi$ , abreviado por  $\log \mathbb{L}_\pi$ -óptimo, si:

$$\begin{aligned} \xi^* &= \arg \min_{\xi} \mathbb{E}_{\pi} [\log (\operatorname{Tr} (K^T(\Theta)M^{-1}(\xi; \Theta)K(\Theta)))] \\ &= \arg \min_{\xi} \int \log (\operatorname{Tr} (K^T(\Theta)M^{-1}(\xi; \Theta)K(\Theta))) \pi(\Theta)d\Theta \end{aligned}$$

Al aplicar el respectivo teorema de equivalencia, Fedorov y Hackl (1997), se tiene que:  $\xi^*$  es un diseño  $\log \mathbb{L}_\pi$ -óptimo si y sólo si:

$$\forall t \in [0, t_{\infty}] \quad \mathbb{E}_{\pi} \left[ \frac{f^T(t)M^{-1}(\xi^*)KK^TM^{-1}(\xi^*)f(t)}{\operatorname{Tr} (K^TM^{-1}(\xi^*)K)} - 1 \right] \leq 0.$$

Se observa que hay al menos dos posibles criterios de optimalidad para  $\mathbb{L}$ -optimalidad usando distribuciones a prioris, y como se vé más adelante los diseños resultantes son diferentes. Lo anterior es debido a la no equivalencia de ambos criterios.

### Definición de las distribuciones a prioris

Con el fin de analizar el efecto que tiene la distribución a priori en el diseño óptimo, para los criterios dados en la sección anterior, se consideran tres familias de distribuciones a prioris. Aquellas familias son un subconjunto de todas las distribuciones discretas uniformes definidas en algún subconjunto apropiado de  $\mathbb{R}^m$ . Todas las familias se supone que varían alrededor de algún valor local de  $\Theta$ ,  $\Theta_0^T = (\theta_{01}, \dots, \theta_{0m})$ .

1. Distribuciones a prioris simétricas, denotada por  $S(\delta)$ . El soporte de cualquier distribución de esta familia contiene a  $\Theta_0$  y todas las combinaciones definidas en el conjunto:

$$\Omega_{I\delta} = \{ \Theta^T = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m : \theta_i = (1 + \delta)\theta_{0i} \text{ ó } \theta_i = (1 - \delta)\theta_{0i} \}.$$

2. Distribuciones a prioris asimétricas a la izquierda, denotada por  $LA(\delta)$ . El soporte de cualquier distribución de esta familia contiene  $\Theta = (1 - \delta/2)\Theta_0$  y todas las combinaciones definidas en el conjunto:

$$\Omega_{II\delta} = \{ \Theta_0 \} \cup \{ \Theta^T = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m : \theta_i = (1 - \delta)\theta_{0i} \}.$$

3. Distribuciones a prioris asimétricas a la derecha, denotada por  $RA(\delta)$ . El soporte de esta familia contiene  $\Theta = (1 + \delta/2)\Theta_0$  y todas las combinaciones definidas sobre:

$$\Omega_{III\delta} = \{ \Theta_0 \} \cup \{ \Theta^T = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m : \theta_i \in (1 + \delta)\theta_{0i} \}.$$

En lo que sigue, una distribución a priori cuyo soporte sea  $\Omega_{I\delta}$  ó  $\Omega_{II\delta}$  ó  $\Omega_{III\delta}$  será denotada por  $S\pi_\delta$ ,  $LA\pi_\delta$  o  $RA\pi_\delta$ , respectivamente.

Intuitivamente en cada una de las familias de distribuciones a prioris definidas antes se están considerando tres casos: El caso donde me equivoco en la especificación del valor  $\Theta_0$  considerando valores posibles para  $\Theta$  más grandes y más pequeños que las componentes de  $\Theta_0$ . El error cometido estaría reflejado en el valor de  $\delta$ .

- En el primero, distribuciones a prioris simétricas, se tiene en cuenta un margen  $\delta$  de error alrededor del valor local  $\Theta_0$ , y luego se consideran las diferentes alternativas para posibles valores de  $\Theta$ , ya sea que la respectiva componente de  $\Theta_0$  quede fija o tome un valor más grande o más pequeño de ésta. En la práctica este sería el caso más común.
- En el segundo, distribuciones a prioris asimétricas a la izquierda, se consideran los casos donde los valores de  $\Theta$  estuvieran por debajo del valor local, valores más pequeños, por una cantidad de  $(1 - \delta)$ . Es decir, con estas distribuciones se hallarían diseños óptimos para una distribución a priori cuyo soporte son diferentes valores de  $\Theta$  que están por debajo del valor local  $\Theta_0$ . En este caso  $\Theta_0$  se considera una sobreestimación de  $\Theta$ .
- En el tercero, distribuciones a prioris asimétricas a la derecha. Este caso es lo contrario del anterior, se tienen distribuciones a prioris con puntos de soporte valores del vector de parámetros  $\Theta$  que están relativamente por encima, más grandes, del valor local  $\Theta_0$ . En este caso  $\Theta_0$  es una subestimación de  $\Theta$ .

### Ejemplo ilustrativo usando distribuciones a prioris

Para el modelo  $I$ , dado por la ecuación 4.5, se hallan los diseños  $\mathbb{D}_\pi$  y  $\mathbb{L}_\pi$ -óptimos promediados por la distribución a priori  $\pi$ , siendo  $\pi$  una distribución a priori simétrica, asimétrica a la izquierda ó asimétrica a la derecha. Los valores para  $\delta$  considerados fueron:  $\{0.05, 0.10, 0.15\}$ . Además, con propósitos ilustrativos y comparativos, se tomó el mismo valor local para  $\Theta$ ,  $\Theta_0^T = (0.30, 0.20, 0.15, 0.05, 0.08, 0.25)$ .

De manera análoga a los diseños  $\mathbb{L}$  y  $\mathbb{D}$ -óptimos se implementó el algoritmo del Apéndice A para  $\phi_\pi$ -optimalidad. Los diseños óptimos obtenidos se encuentran en la Tabla 4.5.

Además, en las Figuras 4.9 y 4.12 se muestran los gráficos de las funciones de las varianzas (sensibilidad), para cada una de las distribuciones a prioris consideradas y para los tres criterios de optimalidad  $\mathbb{D}_\pi$ ,  $\mathbb{L}_\pi$  y  $\log \mathbb{L}_\pi$ .

**Algunos comentarios acerca de los gráficos**

Se observa que los diseños  $\mathbb{D}_\pi$ -óptimos tienen el mismo peso en las diferentes distribuciones a priori consideradas. Además, en las distribuciones simétricas y asimétricas a la izquierda existe una tendencia creciente en la magnitud de los puntos de soporte con respecto a  $\delta$ , ocurre lo contrario en las distribuciones a priori asimétricas a la derecha, siendo razonable tener que muestrear en tiempos más cortos cuando las tasas de transferencia, en general, toman valores más altos, ver Figura 4.10.

En los diseños  $\mathbb{L}_\pi$ -óptimos se observa una tendencia creciente en los pesos de cada uno de los diseños, proporcionando pesos pequeños, menos del 10 %, hasta el tercer tiempo de muestreo, y luego pesos que se van incrementando hasta un 43 %, aproximadamente. También se resalta un comportamiento análogo a los diseños  $\mathbb{D}_\pi$ -óptimos con respecto a los puntos de puntos, es decir, hay un relativo crecimiento de la magnitud de los puntos de soporte con respecto a  $\delta$  en las distribuciones simétricas y asimétricas a la izquierda, y un decrecimiento en las distribuciones asimétricas a la derecha, ver Figura 4.10. Las mismas anotaciones son válidas para el caso de los diseños  $\log \mathbb{L}_\pi$ -óptimos dados en la Figura 4.11 y Tabla 4.6.

Distribución a priori	Diseño óptimo						Valor óptimo
$D_{S\pi_{0.05}}$	1.0809	3.8389	9.0319	18.4800	35.7158	67.8997	20.79
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$D_{S\pi_{0.10}}$	1.0810	3.8405	9.0414	18.5137	35.8037	68.1334	21.01
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$D_{S\pi_{0.15}}$	1.0812	3.8432	9.0575	18.5703	35.9523	68.5291	21.40
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$D_{LA\pi_{0.05}}$	1.1087	3.9370	9.2610	18.9454	36.6096	69.5823	20.44
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$D_{LA\pi_{0.10}}$	1.1379	4.0411	9.5076	19.4539	37.5989	71.4821	20.18
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$D_{LA\pi_{0.15}}$	1.1686	4.1511	9.7698	19.9981	38.6632	73.5419	19.97
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$D_{RA\pi_{0.05}}$	1.0546	3.7449	8.8092	18.0209	34.8230	66.1862	21.03
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$D_{RA\pi_{0.10}}$	1.0295	3.6561	8.6015	17.5990	34.0125	64.6595	21.37
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$D_{RA\pi_{0.15}}$	1.0056	3.5716	8.4045	17.2005	33.2497	63.2309	21.72
	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	
$L_{S\pi_{0.05}}$	0.7201	3.5444	9.0325	19.1103	41.9188	73.8642	$24.98 \times 10^3$
	0.0493	0.0544	0.0809	0.1643	0.2319	0.4192	
$L_{S\pi_{0.10}}$	0.6809	3.4932	9.0195	19.1882	40.6821	77.7939	$48.80 \times 10^3$
	0.0578	0.0565	0.0928	0.1618	0.2356	0.3955	
$L_{S\pi_{0.15}}$	0.6610	3.4775	8.9645	19.1316	39.2367	80.7581	$21.51 \times 10^4$
	0.0669	0.0623	0.1034	0.1674	0.2290	0.3710	
$L_{LA\pi_{0.05}}$	0.7587	3.6664	9.2460	19.5311	43.3292	74.0924	$22.14 \times 10^3$
	0.0465	0.0546	0.0758	0.1672	0.2253	0.4306	
$L_{LA\pi_{0.10}}$	0.7681	3.7509	9.5131	20.1118	44.2231	77.1947	$25.75 \times 10^3$
	0.0478	0.0541	0.0784	0.1645	0.2285	0.4267	
$L_{LA\pi_{0.15}}$	0.7666	3.8227	9.7850	20.7340	44.8797	81.4020	$34.40 \times 10^3$
	0.0512	0.0542	0.0836	0.1610	0.2326	0.4174	
$L_{RA\pi_{0.05}}$	0.7104	3.4658	8.7950	18.5925	41.1184	71.1356	$21.64 \times 10^3$
	0.0481	0.0547	0.0788	0.1664	0.2300	0.4220	
$L_{RA\pi_{0.10}}$	0.6776	3.3597	8.5964	18.2034	39.8251	70.8355	$24.44 \times 10^3$
	0.0508	0.0547	0.0833	0.1643	0.2349	0.4120	
$L_{RA\pi_{0.15}}$	0.6443	3.2576	8.3943	17.8225	38.4039	70.8799	$30.75 \times 10^3$
	0.0546	0.0555	0.0890	0.1631	0.2377	0.4001	

Tabla 4.5: Diseños  $\mathbb{D}_\pi$  y  $\mathbb{L}_\pi$ -óptimos con las distribuciones a prioris, modelo  $I$ .

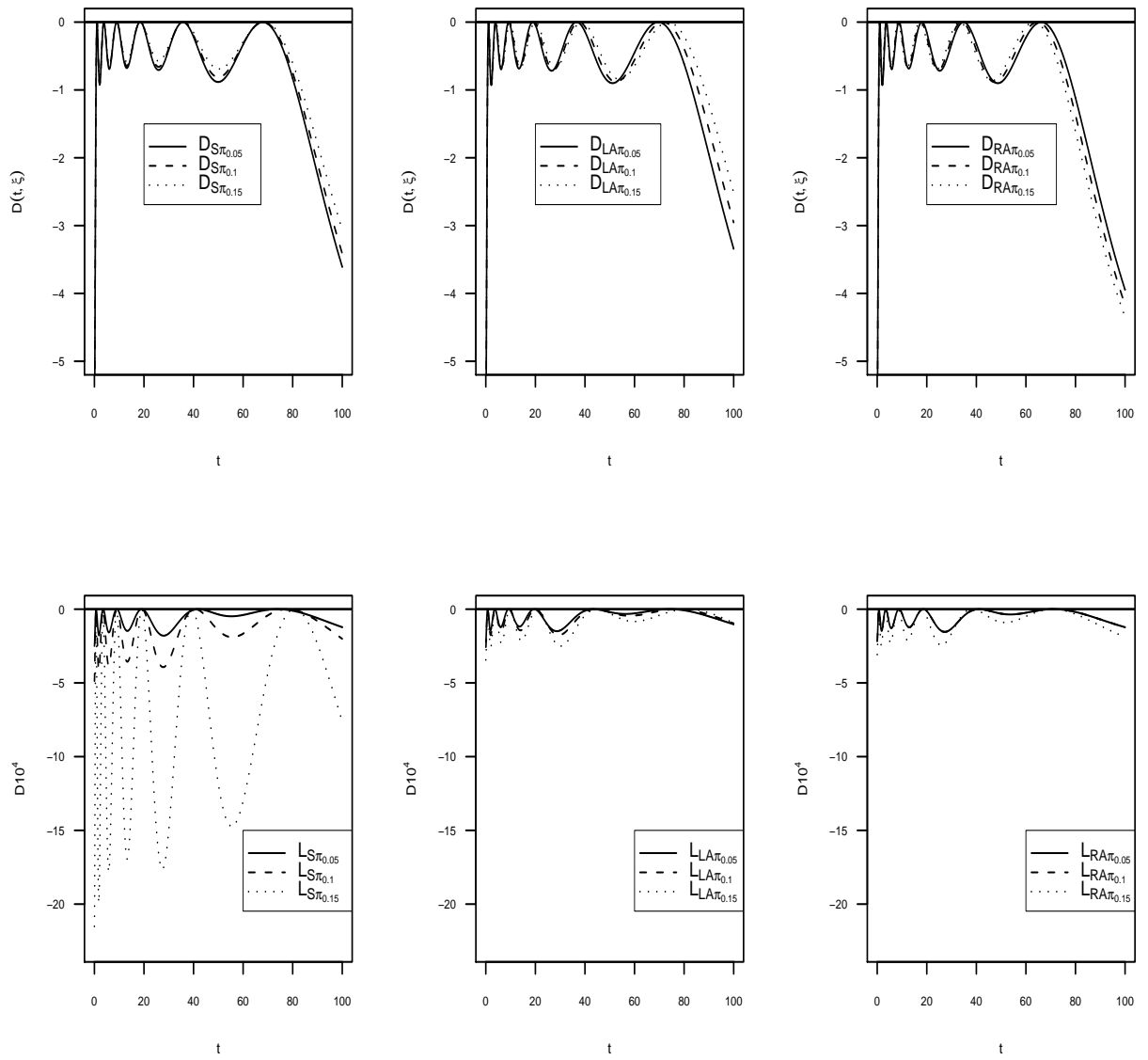


Figura 4.9: Función de varianza conjunto para  $\mathbb{D}_\pi$  y  $\mathbb{L}_\pi$ -optimalidad, modelo  $I$ .



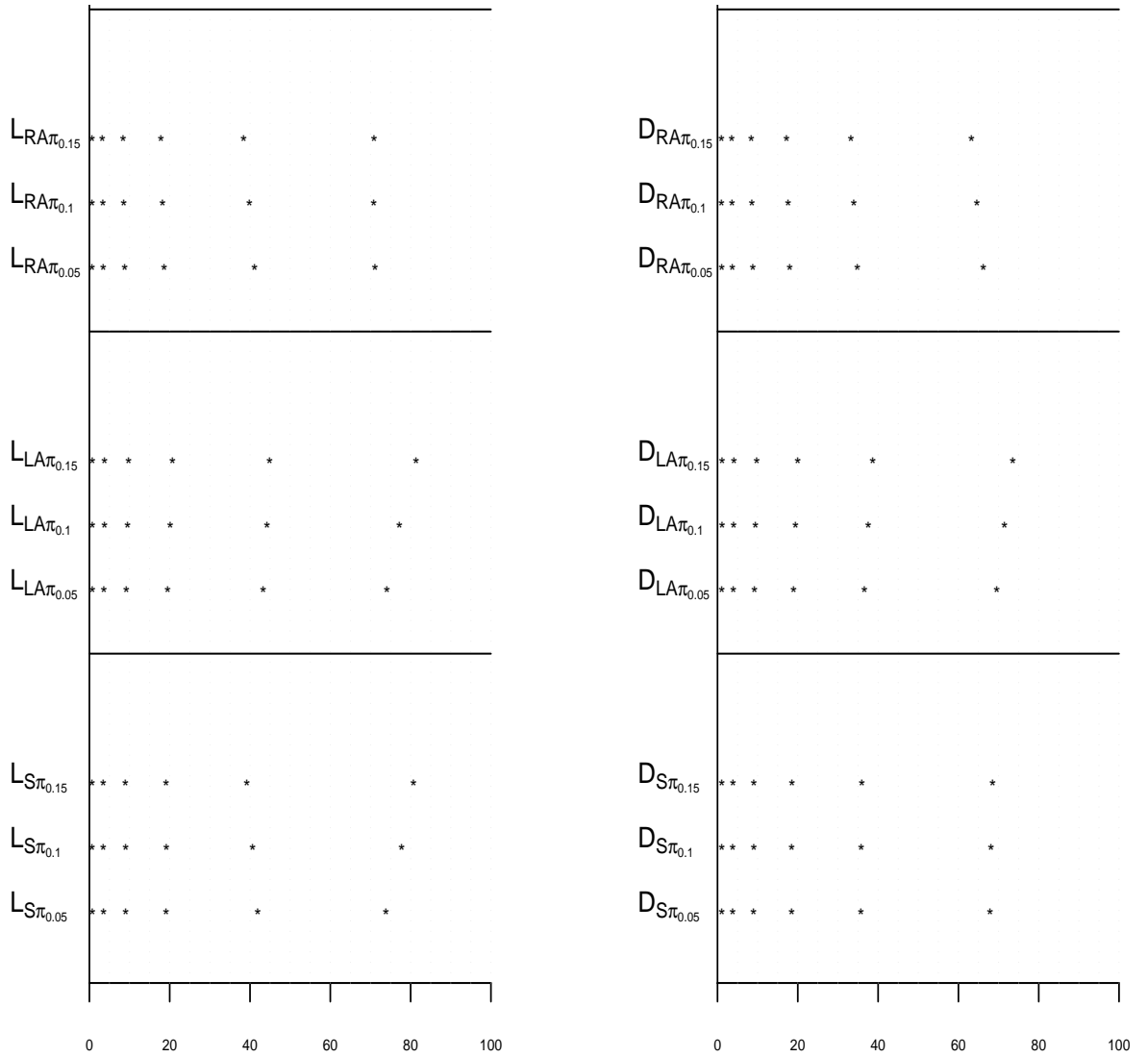


Figura 4.10: Comparación de los puntos de soporte para los diseños  $\mathbb{D}_\pi$  y  $\mathbb{L}_\pi$ -óptimos, modelo  $I$ .

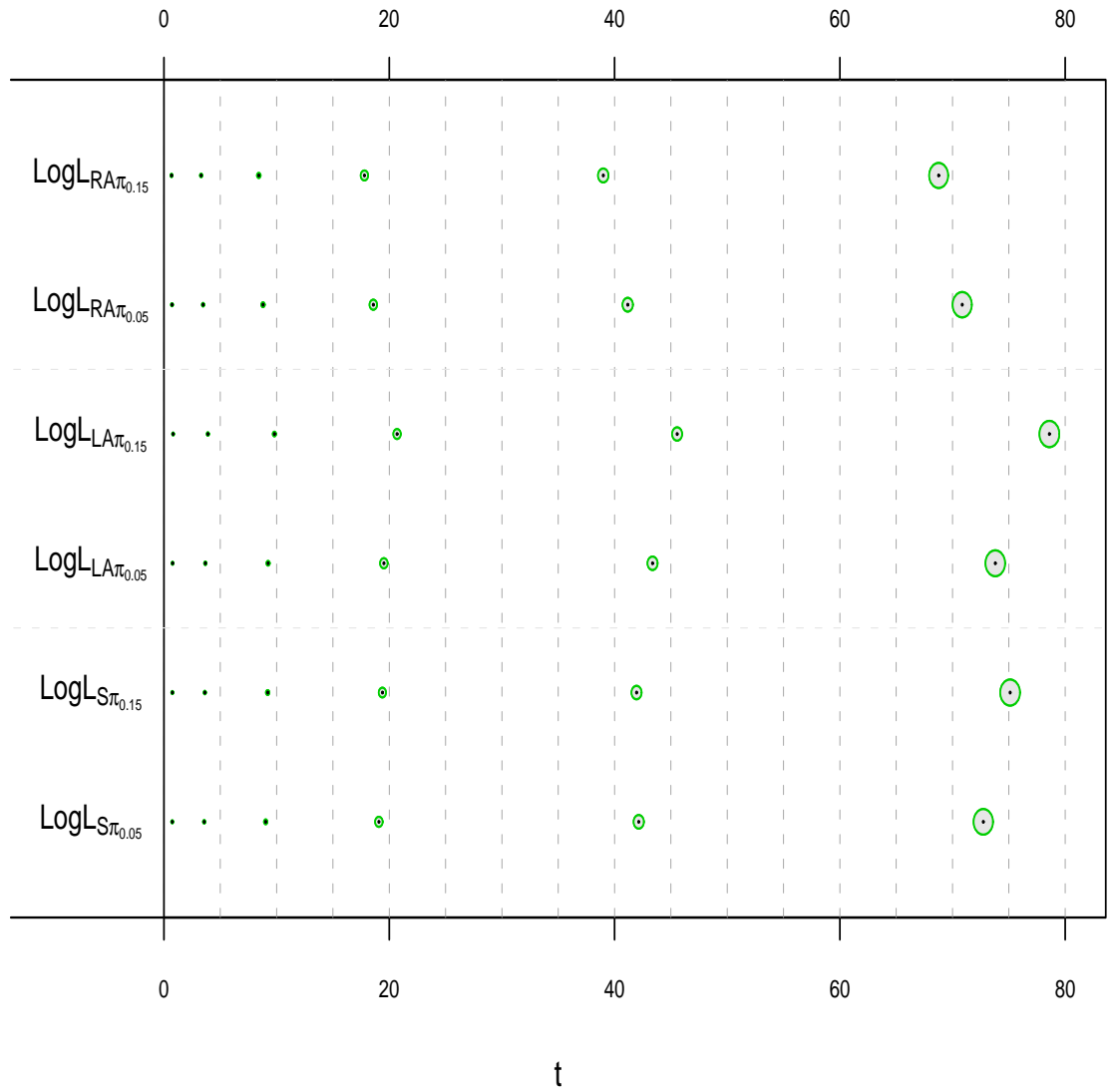


Figura 4.11: Gráfico conjunto tanto de los puntos de soporte ( $t$ ) junto con el peso: proporcional al radio del círculo, para los diseños  $\log \mathbb{L}_\pi$ -óptimos, modelo  $I$ .

Distribución a priori	Diseño óptimo						Valor óptimo
$\log L_{S\pi_{0.05}}$	0.7360	3.5705	9.0308	19.0758	42.1434	72.7289	10.27
	0.0471	0.0544	0.0771	0.1660	0.2268	0.4286	
$\log L_{S\pi_{0.10}}$	0.7363	3.5861	9.0996	19.2022	41.9587	73.8947	10.36
	0.0479	0.0540	0.0782	0.1629	0.2263	0.4307	
$\log L_{S\pi_{0.15}}$	0.7450	3.6245	9.2077	19.3879	41.9497	75.1092	10.84
	0.0480	0.0538	0.0782	0.1603	0.2237	0.4360	
$\log L_{LA\pi_{0.05}}$	0.7629	3.6733	9.2425	19.5184	43.3709	73.7947	9.99
	0.0460	0.0546	0.0749	0.1678	0.2237	0.4331	
$\log L_{LA\pi_{0.10}}$	0.7866	3.7818	9.5076	20.0708	44.4411	75.9640	10.09
	0.0457	0.0543	0.0744	0.1667	0.2223	0.4365	
$\log L_{LA\pi_{0.15}}$	0.8110	3.8988	9.8054	20.6883	45.5517	78.5964	10.25
	0.0457	0.0539	0.0743	0.1649	0.2216	0.4396	
$\log L_{RA\pi_{0.05}}$	0.7141	3.4717	8.7929	18.5819	41.1650	70.8581	9.97
	0.0476	0.0547	0.0778	0.1669	0.2286	0.4244	
$\log L_{RA\pi_{0.10}}$	0.6913	3.3816	8.5977	18.1746	40.0540	69.7923	10.03
	0.0486	0.0546	0.0797	0.1657	0.2306	0.4209	
$\log L_{RA\pi_{0.15}}$	0.6714	3.3009	8.4178	17.7943	38.9938	68.7663	10.14
	0.0495	0.0546	0.0812	0.1644	0.2315	0.4187	

Tabla 4.6: Diseños  $\log \mathbb{L}_\pi$ -óptimos con diferentes distribuciones a prioris, modelo  $I$ .

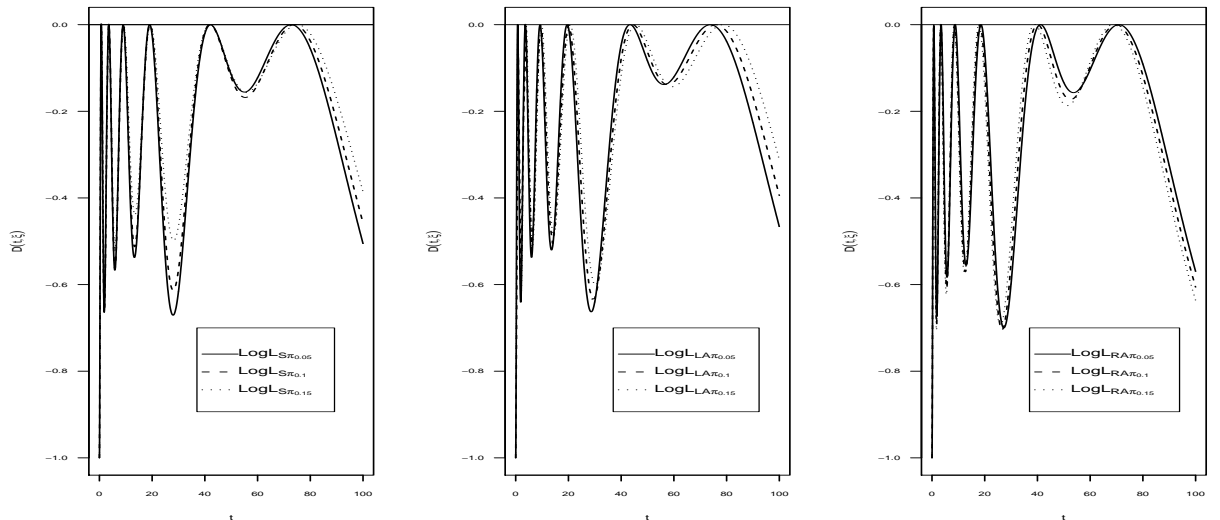


Figura 4.12: Función de varianza conjunto para los diseños  $\log \mathbb{L}_\pi$ -óptimos, modelo  $I$ .

$\Theta$	Diseño óptimo ( $\xi_T$ )	Valor óptimo $\Delta_2(\xi_T)$										
$\Theta_1$	<table border="1"> <tr> <td>0.706</td> <td>3.798</td> <td>10.262</td> <td>24.165</td> <td>61.022</td> </tr> <tr> <td>0.075</td> <td>0.065</td> <td>0.139</td> <td>0.248</td> <td>0.473</td> </tr> </table>	0.706	3.798	10.262	24.165	61.022	0.075	0.065	0.139	0.248	0.473	$2.92 \times 10^{-6}$
0.706	3.798	10.262	24.165	61.022								
0.075	0.065	0.139	0.248	0.473								

 Tabla 4.7: Diseños  $\mathbb{T}$ -óptimos locales usando el valor local  $\Theta_1 = \Theta_0$ .

#### 4.3.4. Diseños óptimos para discriminar entre ambos modelos

##### Diseños $\mathbb{T}$ -óptimos Locales

Si en el modelo de cuatro compartimientos se hace  $\theta_2 = \theta_3 = 0$ , se muestra fácilmente, que la función determinística del modelo  $I$  se reduce a la función respectiva en el modelo  $II$ , obtenido al considerar el modelo de tres compartimientos. Se concluye que ambos modelos están anidados. En esta sección y las siguientes, con el fin de evitar confusiones, se usa  $\Theta_1$  y  $\Theta_2$ , como los vectores de parámetros asociados a los modelos  $I$  y  $II$ , respectivamente; mientras  $\eta_1(t; \Theta_1)$  y  $\eta_2(t; \Theta_2)$ , denotan las funciones determinísticas halladas para la concentración en el compartimiento central en ambos modelos.

El primer modelo, el modelo de seis parámetros, se asume como el modelo verdadero, es decir,  $\eta_i(t; \Theta) = \eta_1(t; \Theta_1)$ , donde se considera como valor local para el vector de parámetros  $\Theta^T = \Theta_1^T = (0.30, 0.10, 0.06, 0.04, 0.08, 0.20)$ .

A continuación se halla un diseño  $\mathbb{T}$ -óptimo local que permite discriminar entre ambos modelos, es decir, se encuentra el diseño  $\xi_T$  tal que maximiza:

$$\Delta_2(\xi) = \min_{\Theta_2} \int_{\mathcal{X}} \{\eta_T(t; \Theta_1) - \eta_C(t; \Theta_2)\}^2 d\xi(t) \quad (4.7)$$

En la Tabla 4.7 se muestra el diseño  $\mathbb{T}$ -óptimo local hallado mediante la implementación del respectivo algoritmo, dado en el Apéndice A.2. Se usa el respectivo teorema de equivalencia y se verifica, con ayuda de la Figura 4.13, que en efecto el diseño es  $\mathbb{T}$ -óptimo local.

Se observa, como se puntualizó en la introducción de este capítulo, que el diseño  $\mathbb{T}$ -óptimo se reduce a considerar cinco tiempos de muestreo. A pesar de que este criterio posee buenas propiedades en cuanto a potencia en términos asintóticos, claramente con cinco puntos no es posible hacer estimación de los parámetros en el modelo  $I$ , que consta de seis parámetros. Note que en este diseño, como ocurrió en el diseño  $\mathbb{L}$ -óptimo local, se da un peso considerable al último tiempo de muestreo, 47.3% aproximadamente; observe que un 72% de todas las observaciones de la muestra se reserva a los últimos dos tiempos (24.16 y 61.02).

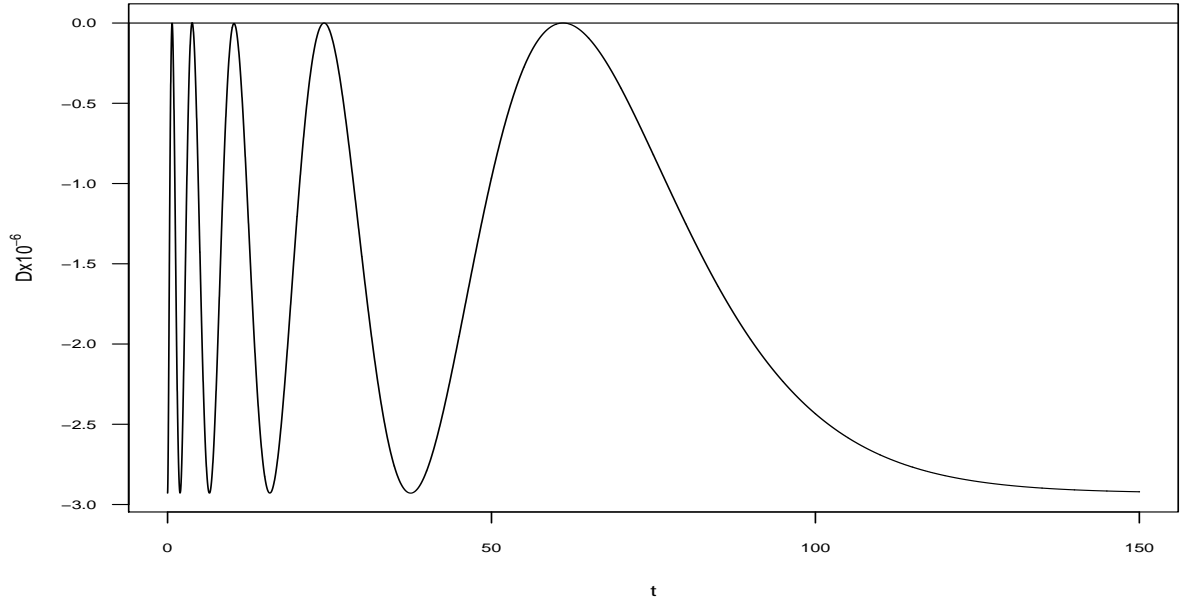


Figura 4.13: Función de varianza para  $\mathbb{T}$ -optimalidad local.

En este caso, como ambos modelos están anidados, es posible considerar otro criterio para discriminar. Consiste en usar un criterio que permita hallar diseños que estimen en forma óptima los parámetros,  $\theta_2$  y  $\theta_3$ . Este criterio se expone a continuación.

#### $\mathbb{D}_v$ -optimalidad para discriminar

En esta sección se considera el criterio  $\mathbb{D}_v$ -optimalidad. Este criterio conduce a diseños que estiman en forma óptima  $v$ -componentes del vector  $\Theta_1$ . Como se anotó antes ambos modelos están anidados. Al sustituir  $\theta_2 = \theta_3 = 0$  en el modelo  $I$ , se obtiene como un caso especial el modelo  $II$ . Una manera de discriminar entre ambos es a partir de la estimación óptima de estos dos parámetros en el modelo  $I$ . Este objetivo igualmente es alcanzado con el criterio  $\mathbb{D}_2$ -optimalidad. Consiste en hallar un diseño  $\xi$  que maximice el determinante de:

$$C_A(\xi; \Theta_1) = (A^T M_1^{-1}(\xi; \Theta_1) A)^{-1},$$

donde  $A_{2 \times 6}^T = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ . Con  $C_A$  la matriz de información asociada a los parámetros  $\theta_2$  y  $\theta_3$ , ver Definición 3.2.2, recuerde que la matriz de información se considera

$\Theta$	Diseño óptimo ( $\xi$ )	$\phi_2(\xi) =  C_A(\xi; \Theta_1) $
$\Theta_1$	$\begin{bmatrix} 0.6906 & 3.6940 & 9.6354 & 19.5259 & 36.2382 & 76.9768 \\ 0.1107 & 0.1005 & 0.1547 & 0.1634 & 0.1822 & 0.2885 \end{bmatrix}$	$3.38 \times 10^{-7}$

 Tabla 4.8: Diseños  $\mathbb{D}_2$ -óptimos locales para discriminación.

una función. Si  $\xi$  es factible para el sistema  $A^T \Theta_1$ , es decir, se cumple la inclusión:

$$\mathcal{C}(A) \subset \mathcal{C}(M_1(\xi; \Theta_1)),$$

entonces el determinante de la matriz de información  $C_A(\xi; \Theta_1)$  se puede expresar como:

$$\det(C_A(\xi; \Theta_1)) = \frac{\det M_1(\xi; \Theta_1)}{\det M_{11}(\xi; \Theta_1)},$$

con  $M_{11}$  la matriz de información de Fisher asociada a  $\Theta_1$  sin considerar  $\theta_2$  y  $\theta_3$ .

El criterio es:

$$\phi_2(\xi) = |C_A(\xi; \Theta_1)|,$$

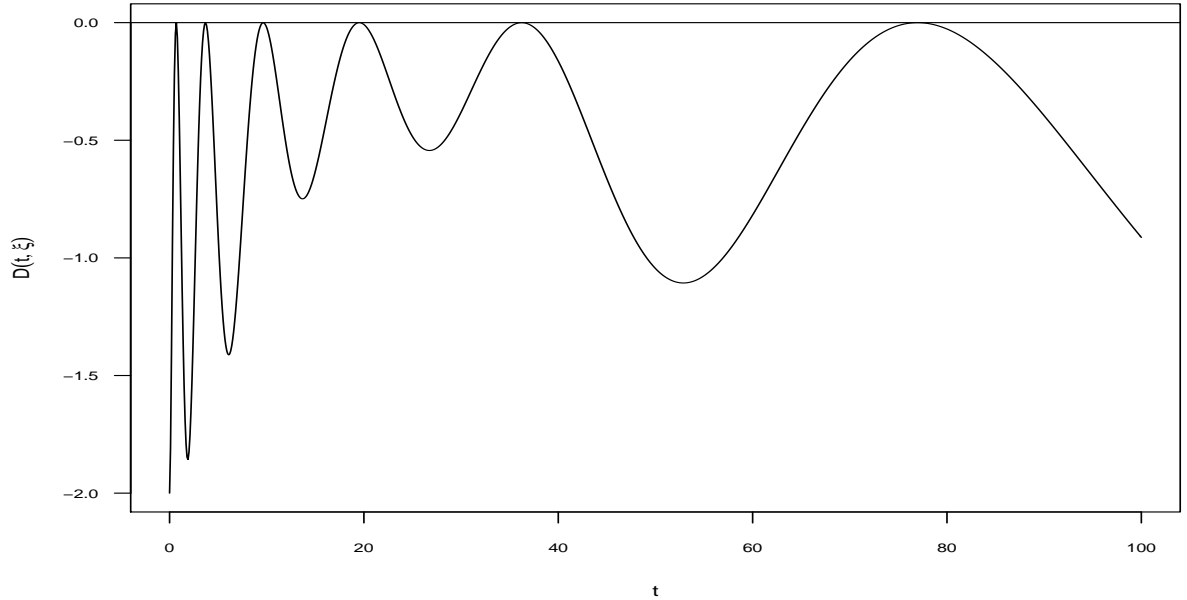
y los diseños  $\mathbb{D}_2$ -óptimos son aquellos que maximizan  $\phi_2(\xi)$ .

Igualmente se halla en forma numérica el diseño óptimo. En la Tabla 4.8 se muestra el diseño  $\mathbb{D}_2$ -óptimo local obtenido con el mismo objetivo de discriminar entre los dos modelos. El diseño consta de seis tiempos de muestreo, a diferencia del  $\mathbb{T}$ -óptimo, éste se puede usar también para la estimación de los parámetros del modelo  $I$ . Este diseño aporta para la estimación de los parámetros del modelo  $I$  una eficiencia del 88.85 %, muy buena comparada con la eficiencia asociada al  $\mathbb{T}$ -óptimo para estimación que es de cero. Además, aproximadamente un 21 % de las observaciones de la muestra se reserva para los dos primeros tiempos, 0.69 y 3.69; mientras un 47 % es para los dos últimos tiempos, 36.24 y 76.98. Hay diferencias de este diseño comparado con el diseño  $\mathbb{T}$ -óptimo, tanto en los tiempos de muestreo como en la proporción de observaciones correspondiente.

Finalmente en la Figura 4.14 está la función de varianza asociada al criterio  $\mathbb{D}_2$ -óptimo. Se verifica que el diseño obtenido es  $\mathbb{D}_2$ -óptimo.

### 4.3.5. Diseños $DEFNS_\alpha$ -óptimos locales

En las secciones previas se han hallado diseños óptimos locales y promediados por distribuciones a priori, usando tanto  $\mathbb{D}$ -optimalidad como  $\mathbb{L}$ -optimalidad, en el modelo  $I$ . Mientras en el modelo  $II$  se hallaron los respectivos diseños óptimos locales. Además


 Figura 4.14: Función de varianza para  $\mathbb{D}_2$ -optimalidad local

se encontraron diseños óptimos para discriminar entre ambos modelos, usando tanto  $\mathbb{T}$ -optimalidad como  $\mathbb{D}_2$ -optimalidad. El énfasis de esta sección es la búsqueda de diseños óptimos que sirvan tanto para discriminar como para estimar las tasas de transferencia en los dos modelos y las funciones no lineales de interés. Además, resultados parciales de esta sección se presentaron en López y Ramos (2008).

Recordando la notación de la Sección 4.2, en ambos modelos, el número de funciones no lineales a estimar es  $l_n = 4$ . Para el  $i$ -ésimo modelo, con  $m_i$  tasas de transferencia ( $m_1 = 6$ ,  $m_2 = 4$ ),  $i = 1, 2$ , las funciones  $H_{ij}$ ,  $j = 1, \dots, 4$ , se calculan a partir de la función  $\eta_i(t; \Theta_i)$ , y representan:

1.  $H_{i1}(\Theta_i)$ , el área bajo la curva de concentración,
2.  $H_{i2}(\Theta_i)$ , el tiempo para la concentración máxima,
3.  $H_{i3}(\Theta_i)$ , la concentración máxima y
4.  $H_{i4}(\Theta_i)$ , el primer tiempo donde la concentración en el modelo  $i$  alcanza el 50 % del valor máximo.

En  $\mathbb{L}$ -optimalidad ó  $\mathbb{D}_{K_i}$ -optimalidad,  $K_i$  denota la matriz de orden  $m_i \times 4$ , con columnas los respectivos gradientes de las funciones  $H_{ij}(\Theta_i)$ .

Inicialmente se construyen diseños  $DEFNS_\alpha$ -óptimos locales para tres valores de  $\alpha$ : 0.25, 0.50 y 0.75. Se usa como criterio para la estimación de las funciones no lineales  $\mathbb{L}$ -optimalidad. Para discriminación se tienen en cuenta los dos criterios mencionados antes y para la estimación de las tasas de transferencia en ambos modelos,  $\mathbb{D}$ -optimalidad. Los criterios se denotan por  $DEFNS_\alpha^{DL}$ -optimalidad y  $DEFNS_\alpha^{TL}$ -optimalidad, donde la letra  $D$ , en este caso, hace referencia al criterio  $\mathbb{D}_2$ -optimalidad. La ecuación 4.2 muestra la expresión a maximizar para el segundo criterio, y en el Teorema 4.2.1 se tiene el teorema de equivalencia. En el caso del primer criterio, se usa el mismo teorema de equivalencia teniendo en cuenta que:

1. El criterio considerado es:

$$\phi_1(\xi) = \left\{ \frac{\det M_1(\xi; \Theta_1)}{\det M_{11}(\xi; \Theta_1)} \right\}^{1/2},$$

donde el exponente es el número de parámetros a estimar.

2. La derivada direccional del  $\log \phi_1(\xi)$  es:

$$D_1(t; \xi) = \frac{1}{2} [f_1^T(t)M_1^{-1}(\xi)f_1(t) - f_{11}^T(t)M_{11}^{-1}(\xi)f_{11}(t)],$$

donde la función  $f_{11}^T(t)$  es el gradiente del vector que resulta de  $\Theta_1$  al omitir  $\theta_2$  y  $\theta_3$ ; y  $M_{11}(\xi)$  es la matriz de información obtenida con la función  $f_{11}(t)$ .

Con lo anterior se tienen las expresiones de los criterios que se van a considerar y su respectivo teorema de equivalencia, ver página 58. La búsqueda de los diseños óptimos es una combinación de varios procedimientos. El primero consiste en encontrar diseños a partir de la adecuación del algoritmo general dado en el Apéndice A con los criterios cóncavos construidos. El segundo procedimiento consiste en la optimización de una función en varias variables -puntos de soporte y pesos- definida apropiadamente, y luego hacer uso de los procedimientos de optimización que ya están programados en varios lenguajes, como por ejemplo, *R* y *Matlab*. En el Apéndice A.3 se muestra un ejemplo genérico programado en *R*. A partir de los ensayos hechos y la verificación mediante el teorema de equivalencia respectivo, se ha podido constatar que el segundo procedimiento, además de rápido, da buenos resultados.

Los diseños óptimos que aparecen en la Tabla 4.9 fueron obtenidos por el segundo procedimiento. Allí se tienen los diseños  $DEFNS_\alpha^{DL}$ -óptimos locales para los valores de  $\alpha$  considerados y los mismos valores locales de los parámetros usados en los modelos



$\alpha$	Diseño óptimo ( $\xi$ )						Valor óptimo
0.25	0.9322	3.7206	9.5987	19.5410	35.9549	67.5850	6.34
	0.1021	0.1123	0.1168	0.1511	0.2262	0.2914	
0.50	0.8229	3.6806	9.6092	19.7227	36.3282	71.1595	6.72
	0.1038	0.1081	0.1330	0.1548	0.2182	0.2821	
0.75	0.7474	3.6810	9.6222	19.6769	36.3610	74.2231	7.09
	0.1070	0.1041	0.1448	0.1593	0.2018	0.2830	

Tabla 4.9: Diseños  $DEFNS_{\alpha}^{DL}$ -óptimos locales.

*I* y *II*. En la Figura 4.15 se tiene el gráfico de la función de varianza presentada en el teorema de equivalencia, es decir, la derivada direccional del criterio en la dirección del diseño unitario en  $t$ . Se verifica que los diseños encontrados son óptimos.

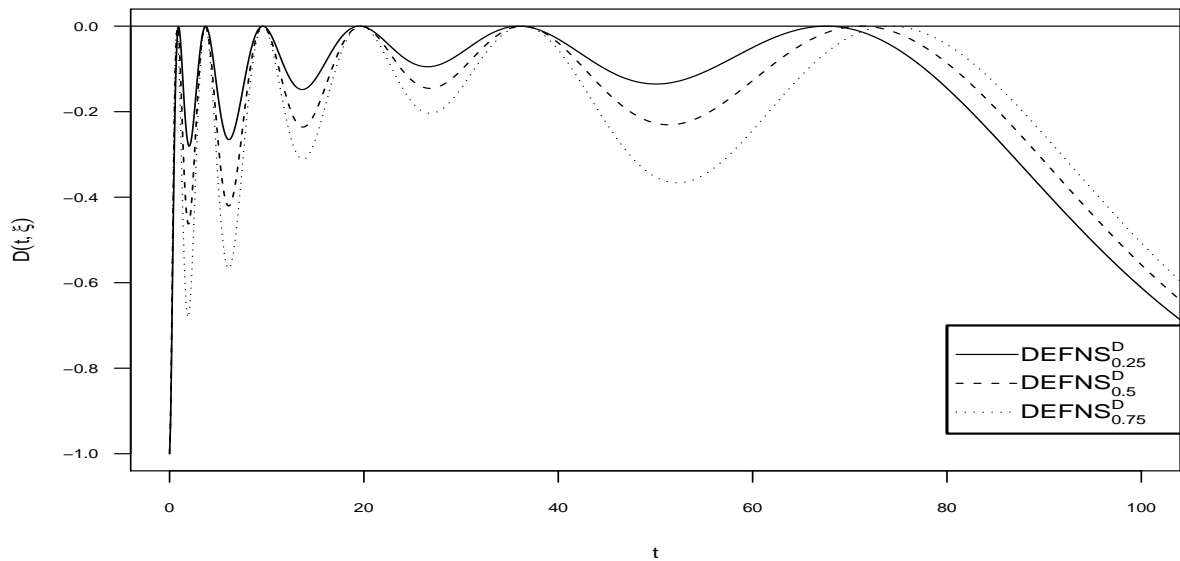


Figura 4.15: Función de varianza para  $DEFNS_{\alpha}^{DL}$ -optimalidad, para  $\alpha = 0.25, 0.50, 0.75$ .

En la Tabla 4.10 se muestran los diseños óptimos obtenidos cuando se usa  $\mathbb{T}$ -optimalidad para discriminar. En la Figura 4.16 se verifica que los respectivos diseños son  $DEFNS_{\alpha}^{TL}$ -óptimos locales.

$\alpha$	Diseño óptimo ( $\xi$ )	Valor óptimo
0.25	0.9783 3.7571 9.9886 21.8765 34.0914 62.6908	7.69
	0.0938 0.1045 0.1173 0.1994 0.0915 0.3935	
0.50	0.8728 3.7070 10.0933 22.8589 35.8963 62.9366	9.41
	0.0862 0.0911 0.1297 0.2262 0.0385 0.4283	
0.75	0.7854 3.7187 10.1958 23.3111 37.7856 63.1089	11.09
	0.0804 0.0779 0.1360 0.2345 0.0156 0.4556	

Tabla 4.10: Diseños  $DEFNS_{\alpha}^{TL}$ -óptimos.

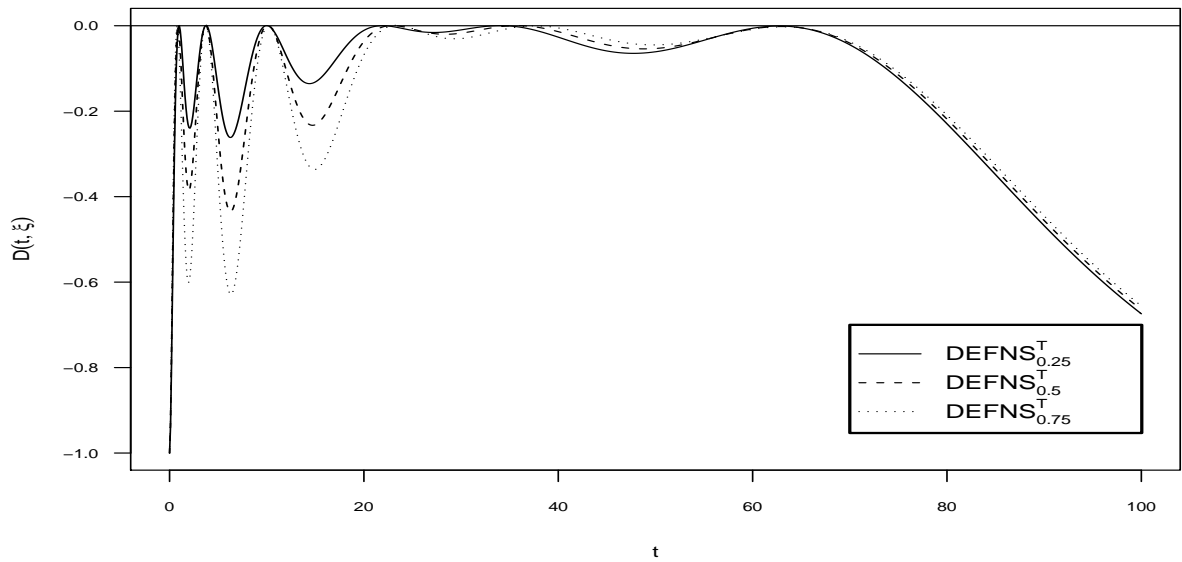


Figura 4.16: Función de varianza para  $DEFNS_{\alpha}^{TL}$ -optimalidad, con  $\alpha = 0.25, 0.50, 0.75$ .

Criterio	Diseño óptimo ( $\xi$ )					
$C_1$	0.82	3.68	9.61	19.72	36.33	71.16
	0.104	0.108	0.133	0.155	0.218	0.282
$C_2$	0.88	3.62	9.09	19.02	35.86	73.08
	0.157	0.161	0.173	0.145	0.158	0.206
$C_3$	0.87	3.71	10.09	22.86	35.90	62.93
	0.086	0.091	0.129	0.226	0.038	0.428
$C_4$	0.92	3.66	9.27	21.92	40.02	60.85
	0.142	0.145	0.162	0.194	0.052	0.305

Tabla 4.11: Diseños  $DEFNS_{0.5}$ -óptimos locales.

### 4.3.6. Comparación entre los diferentes criterios de optimalidad locales

Como se comentó en la sección anterior, a partir de la estructura general del criterio propuesto en la ecuación 4.1- página 57- se pueden obtener dos criterios adicionales. Resultado de considerar  $\mathbb{D}_{K_i}$ -optimalidad para la estimación de las funciones no lineales y los dos criterios usuales para discriminación. En total se tienen cuatro criterios que tienen el mismo objetivo, discriminar y estimar tanto parámetros como las cuatro funciones no lineales de interés. Para esta comparación se tomó  $\alpha = 0.5$ . En la Tabla 4.11 se exhiben los diseños óptimos locales obtenidos al maximizar el respectivo criterio de optimalidad. Se abrevian por:  $C_1 = \phi_\alpha^{D_2L}$ ,  $C_2 = \phi_\alpha^{D_2D_K}$ ,  $C_3 = \phi_\alpha^{TL}$  y  $C_4 = \phi_\alpha^{TD_K}$ .

En los diseños obtenidos usando  $\mathbb{T}$ -optimalidad para discriminar se observa una frecuencia pequeña en el quinto punto de soporte, de hecho son las más pequeñas entre las frecuencias de los cuatro diseños. Se ha observado que este fenómeno ocurre, cuando el valor de  $\alpha$  se acerca a uno, ver por ejemplo los diseños de la Tabla 4.10. Los diseños tienden a aproximarse al diseño de cinco puntos del problema de discriminación entre modelos. Lo anterior está de acuerdo a la forma del criterio propuesto.

Los diseños obtenidos por los cuatro criterios ( $C_1, C_2, C_3, C_4$ ) se compararon con respecto a sus eficiencias, usando los criterios en su versión original, sin la función logaritmo. También se consideraron tres criterios individuales,  $\mathbb{D}$ -optimalidad del modelo  $I$ ,  $\mathbb{D}_2$ - y  $\mathbb{T}$ -optimalidad, denotados por  $C_5 = \phi_D$ ,  $C_6 = \phi_{D_2}$ , y  $C_7 = \phi_T$ , respectivamente. Las eficiencias en porcentaje se encuentran en la Tabla 4.12. En las columnas se tiene el criterio de referencia y en las filas los diseños óptimos obtenidos por cada criterio. Por ejemplo, el diseño  $\xi_1$  tiene un 81.56 % de eficiencia cuando compete con el diseño óptimo obtenido por el tercer criterio ( $\phi_\alpha^{TL}$ ), y en forma análoga se interpretan los demás valores. Se extraen los siguientes hechos relevantes:

$\phi - \text{Ef}(\xi)$ $\times 100\%$	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	Promedio
$\xi_1$	100.00	94.33	81.56	79.11	91.34	97.02	59.98	90.56
$\xi_2$	94.15	100.00	72.31	78.98	97.86	88.15	51.47	88.58
$\xi_3$	67.11	61.45	100.00	94.15	69.21	45.39	93.66	80.93
$\xi_4$	71.12	73.81	93.71	100.00	81.89	52.21	84.23	84.13
$\xi_5$	90.16	96.66	71.66	79.01	100.00	85.34	50.12	87.14
$\xi_6$	97.31	93.38	74.95	73.96	88.85	100.00	55.13	88.08
$\xi_7$	0	0	0	0	0	0	100.00	

Tabla 4.12: Eficiencias de los diseños  $DEFNS_{0.5}$ -óptimos locales,  $C_1 = \phi_{\alpha}^{D_2L}$ ,  $C_2 = \phi_{\alpha}^{D_2DK}$ ,  $C_3 = \phi_{\alpha}^{TL}$ ,  $C_4 = \phi_{\alpha}^{TDK}$ , y los diseños óptimos individuales:  $C_5 = \phi_D$ ,  $C_6 = \phi_{D_2}$  y  $C_7 = \phi_T$ . Donde  $\xi_i$  denota el diseño óptimo obtenido con el criterio  $C_i$ .

- los diseños  $\xi_1$  y  $\xi_4$  tienen buenas eficiencias con respecto al tercer criterio,
- $\xi_2$  presenta eficiencias altas, tanto en el primer criterio como en el quinto; criterio de estimación de los parámetros,
- el diseño  $\mathbb{D}$ -óptimo,  $\xi_5$  tiene una eficiencia de al menos el 50 % para discriminar usando el criterio  $\mathbb{T}$ -óptimo.

Si la decisión para la elección de uno de los criterios estuviera basada en que tan eficiente es el diseño obtenido para estimar los parámetros del modelo y además discriminar entre ambos modelos, entonces los criterios  $C_1$  y  $C_2$  representan una muy buena alternativa. Aunque, el diseño  $\xi_3$  presenta eficiencias altas en los criterios  $C_7$ , y  $C_4$ , donde se usa  $\mathbb{T}$ -optimalidad para discriminar y  $\mathbb{L}$ -optimalidad para estimar las funciones no lineales. En general, no es posible elegir un criterio que cumpla alguna propiedad de robustez con respecto a la eficiencia, es decir, que el diseño óptimo construido por el respectivo criterio tenga eficiencias altas en todos los otros criterios considerados. En la última columna está el promedio de las eficiencias para cada uno de los diseños sin considerar la eficiencia más baja. Se observa una eficiencia promedio alta, por encima del 88 % para los dos primeros diseños óptimos obtenidos con los respectivos criterios. En la siguiente subsección se presenta una metodología general para obtener un diseño que cumpla con las restricciones de tener buenas eficiencias en los demás criterios bajo consideración.

### Metodología para la construcción de un diseño consensual

En esta sección se va a exponer una metodología para hallar un diseño híbrido entre varios diseños óptimos obtenidos por diferentes criterios de optimalidad, pero que satisfacen los mismos objetivos, en nuestro caso discriminación y estimación tanto de

los parámetros como de las funciones no lineales, bajo consideración. Suponga que se tienen  $c$ -criterios de optimalidad, y una ponderación apropiada para cada uno de los criterios  $w_i$   $i = 1, \dots, c$ .

- Primero se calculan los nuevos pesos del diseño al multiplicar los pesos de cada diseño,  $\xi_i$  por su respectiva ponderación,  $w_i$ .
- Luego se proyectan los diferentes puntos de soporte. Se obtienen  $s$ -grupos, donde  $s$ : es el número máximo de puntos de soporte de todos los diseños.
- Para cada grupo,  $g$ , se propone lo siguiente:
  - Cada uno de los puntos de soporte se sustituye por el promedio (o mediana) de las observaciones del grupo, si estos caen en el siguiente intervalo:

$$\text{mediana}_g \pm r_g,$$

donde  $r_g = \lambda(Q_{0.75} - Q_{0.25})/2$ , donde  $Q_r$ : es el cuantil de orden  $r$  asociado al grupo  $g$ . El peso asociado al punto es la suma de los pesos asociados a cada una de las observaciones que están en el intervalo anterior.

- Si la observación (pto de soporte) del grupo  $g$  no está en el intervalo anterior, entonces se conserva esta observación como un punto de soporte adicional, para el nuevo diseño, con su respectivo peso.

El método anterior depende del valor de  $\lambda$ . Dependiendo del valor de  $\lambda$  se pueden obtener diseños, denominados diseños híbridos, de más de  $m$ -puntos de soporte, posiblemente con eficiencias competitivas.

A continuación se presentan dos posibles diseños obtenidos con dos valores de  $\lambda$ , junto con las eficiencias asociadas para los criterios considerados. Se usó como ponderaciones para cada diseño óptimo el promedio de la última columna de la Tabla 4.12. Se tomaron dos valores para  $\lambda$ ,  $\lambda_1 = 3$  y  $\lambda_2 = 3.5$ , dando dos diseños híbridos, de nueve y siete puntos de soporte respectivamente:

$$\xi_{\lambda_1} = \begin{bmatrix} 0.82 & 0.87 & 3.62 & 3.67 & 9.52 & 20.88 & 37.03 & 40.02 & 67.0 \\ 0.02 & 0.09 & 0.04 & 0.09 & 0.15 & 0.18 & 0.11 & 0.01 & 0.30 \end{bmatrix},$$

$$\xi_{\lambda_2} = \begin{bmatrix} 0.88 & 3.67 & 9.52 & 20.88 & 37.03 & 40.02 & 67.0 \\ 0.12 & 0.13 & 0.15 & 0.18 & 0.11 & 0.01 & 0.30 \end{bmatrix}.$$

En la Tabla 4.13 se dan las eficiencias para los dos diseños híbridos. Con ambos diseños se obtienen eficiencias por encima del 85% para todos los criterios excepto para el criterio de discriminación usando  $\mathbb{T}$ -optimalidad, donde se obtuvieron eficiencias alrededor del 74%, mejores que aquellas obtenidas por los diseños asociados a los dos primeros criterios de optimalidad.

%	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
$\xi_{\lambda_1}$	93.66	85.03	90.66	92.25	91.63	85.32	74.31
$\xi_{\lambda_2}$	93.65	85.18	90.66	92.42	91.74	85.25	74.25

Tabla 4.13: Eficiencias de los diseños híbridos  $\xi_{\lambda_j}$ , con respecto a los criterios  $C_1 = \phi_\alpha^{D_2L}$ ,  $C_2 = \phi_\alpha^{D_2D_K}$ ,  $C_3 = \phi_\alpha^{TL}$ ,  $C_4 = \phi_\alpha^{TD_K}$ , y los diseños óptimos individuales:  $C_5 = \phi_D$ ,  $C_6 = \phi_{D_2}$  y  $C_7 = \phi_T$ .

### 4.3.7. Diseños $DEFNS_\alpha^\pi$ -óptimalidad Bayesiano

En esta sección se hallan diseños  $DEFNS_\alpha$ -óptimos promediados por una distribución a priori,  $\pi$ , para  $\Theta^{0T} = (\Theta_1^{0T}, \Theta_2^{0T})$ . Sólo se considera  $\mathbb{L}$ -óptimalidad para la estimación de las funciones no lineales. Con el fin de ver el efecto que tiene la distribución a priori en los diseños óptimos obtenidos, se consideran dos distribuciones a prioris definidas a partir de los valores locales de  $\Theta_1$  y  $\Theta_2$ .

- La primera distribución a priori,  $\pi_1$ , tiene como soporte el conjunto  $\Omega_1 \times \{\Theta_2\}$ , donde  $\Omega_1$  es el soporte asociado a las distribuciones a prioris (simétricas, asimétricas a la izquierda ó asimétricas a la derecha), definidas en la sección 4.3.3, para el valor local de  $\Theta_1$ . El criterio de optimalidad asociado será denotado por:  $DEFNS_{\alpha, S_\delta}^{\otimes, \pi_1}$  ó  $DEFNS_{\alpha, LA_\delta}^{\otimes, \pi_1}$  ó  $DEFNS_{\alpha, RA_\delta}^{\otimes, \pi_1}$ , donde  $\otimes$  es  $D$  ó  $T$ , indicando el criterio usado para discriminar:  $\mathbb{D}_2$  ó  $\mathbb{T}$ -óptimalidad, respectivamente. Los últimos cuatro parámetros de los puntos de soporte de  $\pi_1$  son los valores locales usados para  $\Theta_2$ , mientras que los primeros seis puntos varían dependiendo de la distribución a priori usada. Todas las distribuciones a prioris tienen 65 puntos de soporte, con igual probabilidad.

Los diseños óptimos obtenidos se encuentran en la Tabla 4.16, para valores de  $\delta \in \{0.05, 0.10, 0.15\}$ , y usando  $\mathbb{D}_2$ -óptimalidad para discriminar. En la Figura 4.17 se muestran los gráficos de la función varianza. Se concluye que todos los diseños encontrados son óptimos. Por otro lado, en la Tabla 4.15 se muestran los diseños  $DEFNS_{0.5}^{\pi_1}$  cuando se usa  $\mathbb{T}$ -óptimalidad para discriminación. Usando el respectivo teorema de equivalencia se verificó la optimalidad de todos los diseños obtenidos anteriormente, ver Figura 4.18.

En la Figura 4.19, se comparan todos los diseños óptimos obtenidos usando la distribución a priori  $\pi_1$ , y los dos criterios que se están considerando para discriminación. También se comparan con el diseño óptimo local respectivo, primera fila de cada gráfico. El radio del círculo es proporcional al peso de cada punto de soporte.

Distribución a priori	Diseño Óptimo						Valor óptimo
$DEFNS_{0.5, S_{0.05}}^{D, \pi_1}$	0.8237	3.6828	9.6088	19.7304	36.3794	71.2384	6.74
	0.1039	0.1081	0.1328	0.1552	0.2181	0.2819	
$DEFNS_{0.5, S_{0.10}}^{D, \pi_1}$	0.8258	3.6892	9.6083	19.7562	36.5361	71.4891	6.81
	0.1040	0.1082	0.1323	0.1563	0.2177	0.2815	
$DEFNS_{0.5, S_{0.15}}^{D, \pi_1}$	0.8293	3.6995	9.6101	19.8077	36.8088	71.9452	6.92
	0.1041	0.1084	0.1315	0.1581	0.2172	0.2807	
$DEFNS_{0.5, LA_{0.05}}^{D, \pi_1}$	0.8420	3.7647	9.8564	20.2478	37.1735	73.0529	6.71
	0.1043	0.1082	0.1332	0.1567	0.2179	0.2797	
$DEFNS_{0.5, LA_{0.10}}^{D, \pi_1}$	0.8624	3.8537	10.1148	20.8020	38.0922	75.0976	6.70
	0.1049	0.1084	0.1333	0.1589	0.2173	0.2772	
$DEFNS_{0.5, LA_{0.15}}^{D, \pi_1}$	0.8843	3.9482	10.3853	21.3885	39.0965	77.3167	6.71
	0.1054	0.1086	0.1333	0.1616	0.2163	0.2748	
$DEFNS_{0.5, RA_{0.05}}^{D, \pi_1}$	0.8050	3.6009	9.3723	19.2244	35.5466	69.4005	6.75
	0.1034	0.1079	0.1327	0.1533	0.2182	0.2846	
$DEFNS_{0.5, RA_{0.10}}^{D, \pi_1}$	0.7882	3.5252	9.1451	18.7511	34.8207	67.7629	6.79
	0.1030	0.1078	0.1323	0.1520	0.2179	0.2870	
$DEFNS_{0.5, RA_{0.15}}^{D, \pi_1}$	0.7723	3.4529	8.9272	18.3017	34.1440	66.2362	6.83
	0.1025	0.1076	0.1320	0.1511	0.2173	0.2894	

Tabla 4.14: Diseños  $DEFNS_{\alpha, S_{\delta}}^{D, \pi_1}$  o  $DEFNS_{\alpha, LA_{\delta}}^{D, \pi_1}$  ó  $DEFNS_{\alpha, RA_{\delta}}^{D, \pi_1}$ -óptimos con las distribuciones a priori consideradas,  $\alpha = 0.5$ , y  $\delta = 0.05, 0.10, 0.15$ .

Distribución a priori	Diseño Óptimo						Valor óptimo
$DEFNS_{0.5, S_{0.05}}^{T, \pi_1}$	0.8746	3.7180	10.1025	22.9246	36.8459	63.0659	9.43
	0.0864	0.0913	0.1295	0.2260	0.0433	0.4235	
$DEFNS_{0.5, S_{0.10}}^{T, \pi_1}$	0.8799	3.7469	10.1441	23.0687	38.7007	63.4674	9.51
	0.0868	0.0916	0.1295	0.2253	0.0560	0.4108	
$DEFNS_{0.5, S_{0.15}}^{T, \pi_1}$	0.8884	3.7879	10.2187	23.2271	40.4404	64.2064	9.65
	0.0874	0.0923	0.1296	0.2236	0.0758	0.3913	
$DEFNS_{0.5, LA_{0.05}}^{T, \pi_1}$	0.8921	3.7884	10.3523	23.4674	37.0700	64.5746	9.42
	0.0867	0.0913	0.1297	0.2288	0.0399	0.4236	
$DEFNS_{0.5, LA_{0.10}}^{T, \pi_1}$	0.9133	3.8797	10.6301	24.1389	38.7999	66.3838	9.44
	0.0872	0.0915	0.1298	0.2315	0.0439	0.4161	
$DEFNS_{0.5, LA_{0.15}}^{T, \pi_1}$	0.9370	3.9818	10.9337	24.8672	40.9227	68.3988	9.49
	0.0879	0.0918	0.1299	0.2340	0.0503	0.4061	
$DEFNS_{0.5, RA_{0.05}}^{T, \pi_1}$	0.8552	3.6341	9.8483	22.3119	35.2965	61.4456	9.41
	0.0859	0.0910	0.1295	0.2236	0.0394	0.4305	
$DEFNS_{0.5, RA_{0.10}}^{T, \pi_1}$	0.8390	3.5680	9.6173	21.8140	35.0739	60.0825	9.43
	0.0856	0.0910	0.1293	0.2212	0.0423	0.4306	
$DEFNS_{0.5, RA_{0.15}}^{T, \pi_1}$	0.8240	3.5071	9.4026	21.3610	35.2077	58.8510	9.45
	0.0853	0.0909	0.1292	0.2192	0.0465	0.4288	

Tabla 4.15: Diseños  $DEFNS_{\alpha, S_{\delta}}^{T, \pi_1}$  ó  $DEFNS_{\alpha, LA_{\delta}}^{T, \pi_1}$  ó  $DEFNS_{\alpha, RA_{\delta}}^{T, \pi_1}$ -óptimos para las distribuciones a prioris consideradas,  $\alpha = 0.5$  y  $\delta = 0.05, 0.10, 0.15$ .



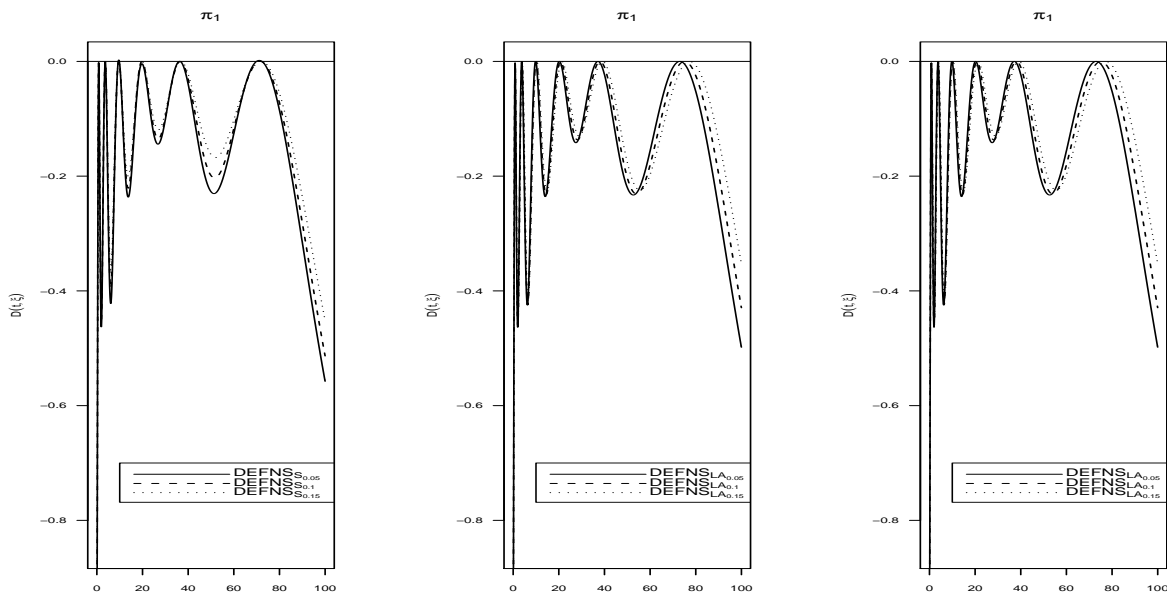


Figura 4.17: Gráfico conjunto de las funciones de varianza para los diseños  $DEFNS_{0.5}$ -óptimos bayesianos considerando la distribución a priori,  $\pi_1$  y criterio de discriminación  $\mathbb{D}_2$ -optimalidad, y de estimación funciones no lineales  $\mathbb{L}$ -optimalidad.

En general se observa, Figura 4.19 (b), la asignación de un peso muy pequeño, sin importar la distribución a priori, en el quinto punto de soporte de cada uno de los diseños que usan  $\mathbb{T}$ -optimalidad para discriminar. Además, en general, en ambos criterios y considerando las distribuciones a prioris asimétricas a la izquierda del valor local, los puntos de los diseños óptimos crecen conforme el parámetro  $\delta$  lo hace. Ocurre lo contrario, en los diseños de las distribuciones a prioris a la derecha del valor local, es decir, los puntos de soporte decrecen si  $\delta$  aumenta.

En las distribuciones simétricas alrededor del valor local y  $\mathbb{D}_2$ -optimalidad, ver Figura 4.19(a), no hay aparente variación de los puntos de soporte a medida que  $\delta$  crece. Hay una gran similitud con el diseño óptimo local. También en la misma figura se observa una tendencia creciente en los pesos de cada uno de los puntos de soporte. Lo anterior no es válido para  $\mathbb{T}$ -optimalidad, Figura 4.19(b), allí se observa que la mayor parte del peso de cada diseño se concentra en el último punto de soporte, entre el 40 % y el 43 %, ver Tabla 4.15. Los diseños simétricos, igualmente presentan gran similitud con el diseño óptimo local respectivo.

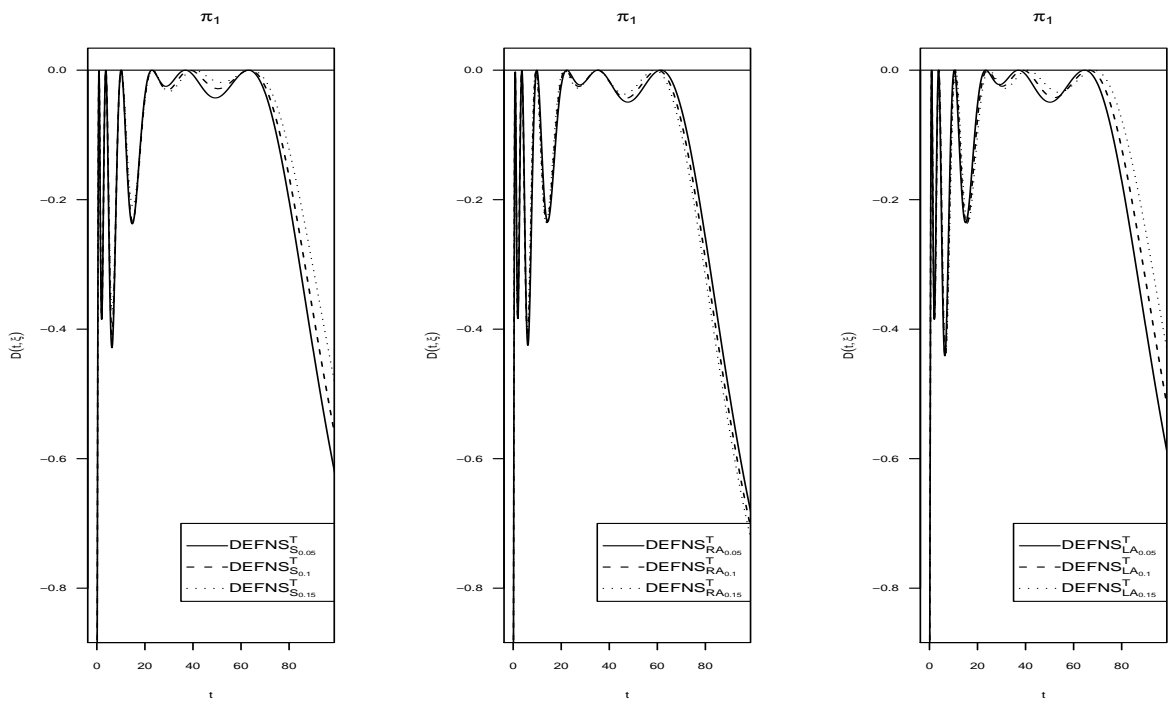


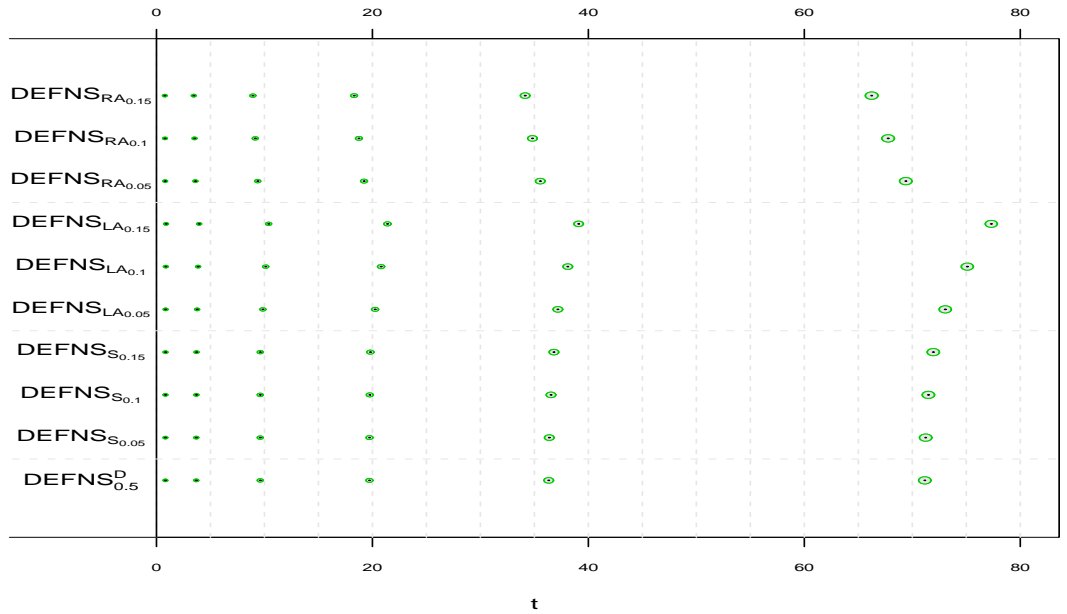
Figura 4.18: Gráfico conjunto de las funciones de varianza para los diseños  $DEFNS_{0.5}$ -óptimos bayesianos considerando la distribución a priori,  $\pi_1$ , usando como criterio de discriminación T-optimalidad, y para estimación de las funciones no lineales L-optimalidad.

Distribución a priori	Diseño óptimo						Valor óptimo
$DEFNS_{0.5, S_{0.05}}^{D, \pi_2}$	0.8236	3.6828	9.6080	19.7275	36.3852	71.2456	6.75
	0.1039	0.1081	0.1328	0.1552	0.2177	0.2824	
$DEFNS_{0.5, S_{0.10}}^{D, \pi_2}$	0.8256	3.6891	9.6051	19.7448	36.5597	71.5189	6.81
	0.1040	0.1082	0.1323	0.1563	0.2163	0.2830	
$DEFNS_{0.5, S_{0.15}}^{D, \pi_2}$	0.8288	3.6989	9.6026	19.7826	36.8635	72.0159	6.92
	0.1041	0.1083	0.1315	0.1580	0.2142	0.2839	
$DEFNS_{0.5, LA_{0.05}}^{D, \pi_2}$	0.8442	3.7756	9.8553	20.2297	37.2750	73.0070	6.93
	0.1039	0.1081	0.1329	0.1549	0.2181	0.2822	
$DEFNS_{0.5, LA_{0.10}}^{D, \pi_2}$	0.8670	3.8769	10.1136	20.7664	38.3068	75.0051	6.70
	0.1039	0.1081	0.1328	0.1552	0.2177	0.2824	
$DEFNS_{0.5, LA_{0.15}}^{D, \pi_2}$	0.8915	3.9852	10.3852	21.3367	39.4372	77.1797	6.70
	0.1039	0.1081	0.1325	0.1558	0.2169	0.2827	
$DEFNS_{0.5, RA_{0.05}}^{D, \pi_2}$	0.8030	3.5913	9.3745	19.2427	35.4553	69.4437	6.75
	0.1039	0.1081	0.1329	0.1549	0.2181	0.2822	
$DEFNS_{0.5, RA_{0.10}}^{D, \pi_2}$	0.7844	3.5072	9.1505	18.7876	34.6475	67.8453	6.79
	0.1039	0.1081	0.1328	0.1551	0.2178	0.2823	
$DEFNS_{0.5, LA_{0.15}}^{D, \pi_2}$	0.7668	3.4276	8.9366	18.3557	33.8874	66.3531	6.84
	0.1039	0.1081	0.1326	0.1555	0.2173	0.2826	

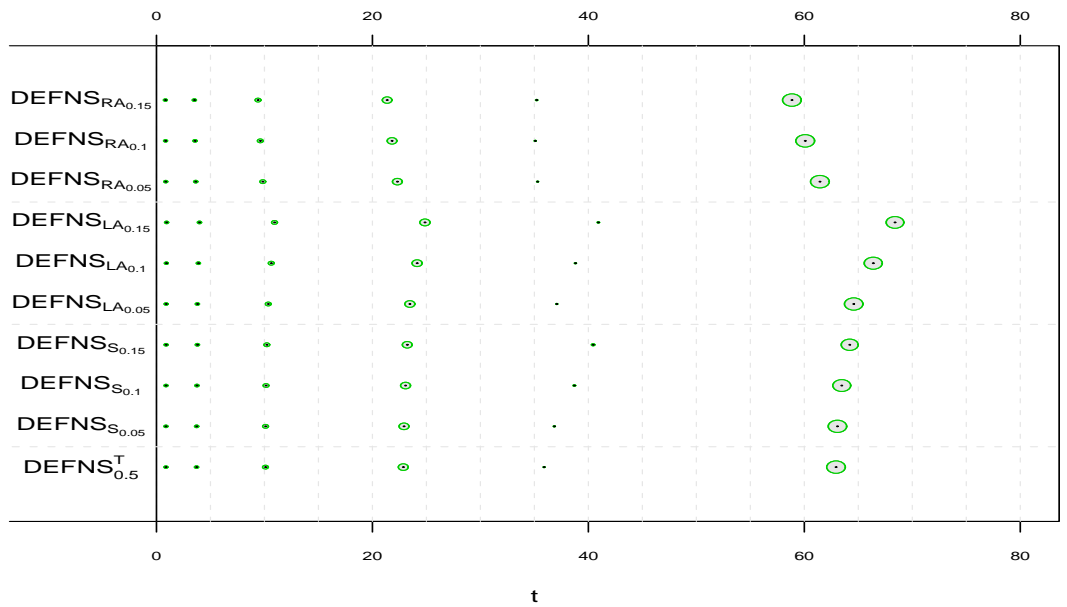
Tabla 4.16: Diseños  $DEFNS_{\alpha, S_\delta}^{D, \pi_2}$  ó  $DEFNS_{\alpha, LA_\delta}^{D, \pi_2}$  ó  $DEFNS_{\alpha, RA_\delta}^{D, \pi_2}$ -óptimos para las distribuciones a prioris consideradas,  $\alpha = 0.5$  y  $\delta = 0.05, 0.10, 0.15$ .

En el criterio de optimalidad donde se usa  $\mathbb{D}_2$ -optimalidad para discriminar se concluye que los diseños óptimos bayesianos con distribuciones simétricas alrededor de un valor local presentan valores muy similares al diseño óptimo local respectivo. Esto no ocurre con  $\mathbb{T}$ -optimalidad, donde el quinto punto de soporte de las distribuciones a prioris simétrica, crece conforme  $\delta$  crece. El diseño óptimo local tiende a hacer muy parecido cuando hay poca incertidumbre alrededor del valor local ( $\delta = 0.05$ ) en las distribuciones a prioris simétricas.

- La segunda distribución a priori considerada es,  $\pi_2$ . Tiene soporte en  $\Omega_{\pi_2}$ : el conjunto de todos los puntos de soporte asociados con las distribuciones a prioris definidas en la sección 4.3.3, donde  $\Theta^{0T} = (\Theta_1^{0T}, \Theta_2^{0T})$ . Se usa la misma notación de la distribución a priori  $\pi_1$  para denotar los diseños obtenidos. Las distribuciones a prioris resultantes tienen un total de 1025 puntos de soporte con igual probabilidad. Los diseños óptimos promediados por la distribución a priori  $\pi_2$  se hallan en la Tabla 4.16, para el caso de  $\mathbb{D}_2$ -optimalidad para discriminación. La optimalidad de todos diseños anteriores se verifica mediante la aplicación del respectivo teorema de equivalencia.



(a)



(b)

Figura 4.19: Gráfico conjunto de los puntos de soporte y los pesos cuando se usa para discriminación: (a)  $\mathbb{D}_2$ -optimalidad, (b)  $\mathbb{T}$ -optimalidad. Para la estimación de las funciones no lineales:  $\mathbb{L}$ -optimalidad. En la parte inferior diseño óptimo local, respectivo

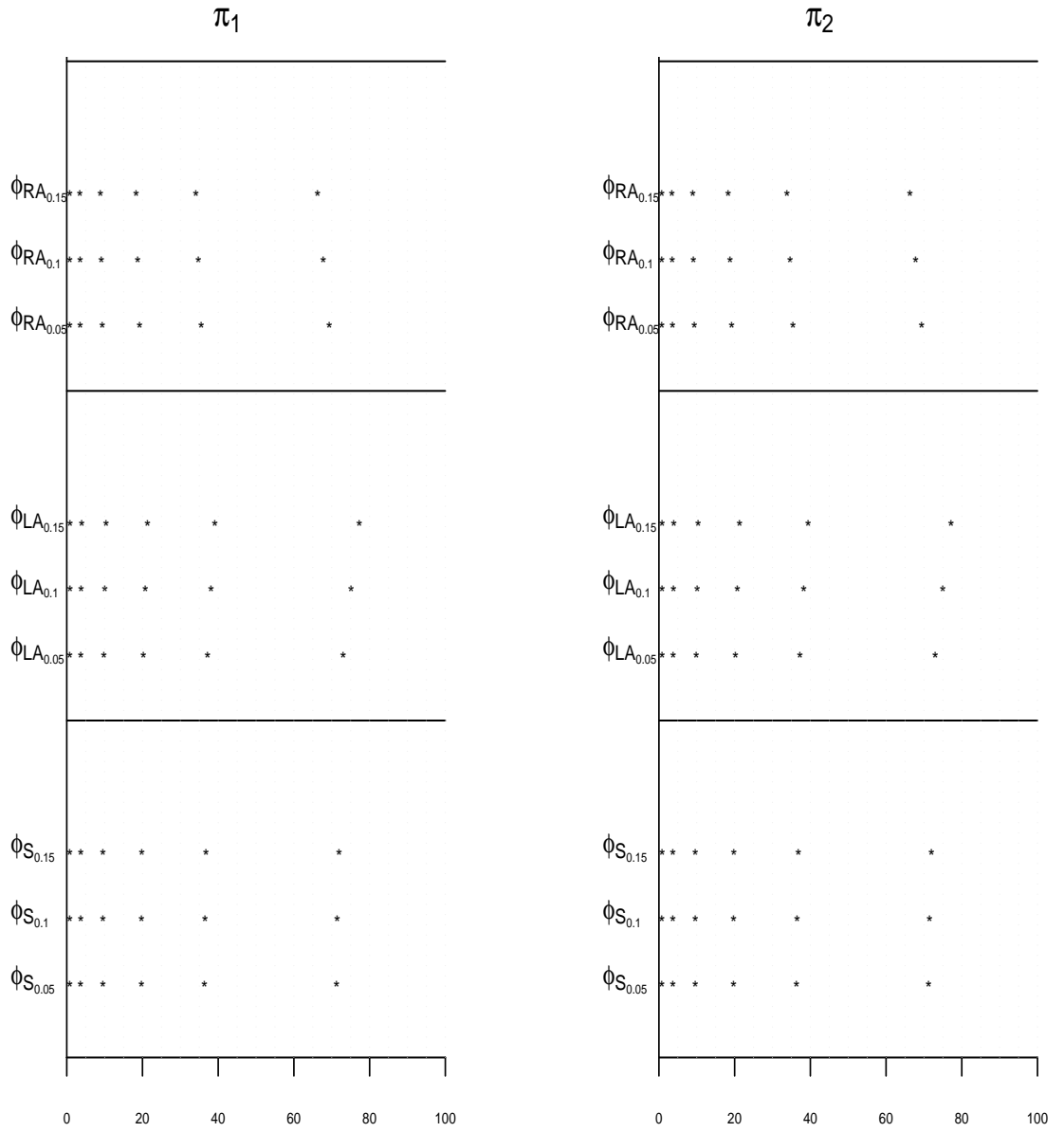


Figura 4.20: Gráfico conjunto de los puntos de soporte para los diseños  $DEFNS_\alpha$ -óptimos ( $\phi_\alpha$ ) considerando ambas distribuciones a prioris,  $\pi_1$  y  $\pi_2$ , y  $\mathbb{D}_2$ -óptimalidad,  $\alpha = 0.5$ .

La Figura 4.20 muestra los puntos de soporte asociados a los diseños óptimos de todas las distribuciones a prioris con  $\pi_1$  y  $\pi_2$ . En ambos casos hay muy poca diferencia en los puntos de soporte. Además, en esta figura, se observa una tendencia de los dos últimos puntos de soporte, de los diseños asimétricos a la izquierda, de tomar valores más grandes conforme  $\delta$  aumenta. En tanto que ocurre lo contrario con los dos últimos puntos de soporte de aquellos diseños obtenidos con las distribuciones a prioris a la derecha. En los diseños obtenidos con las distribuciones a prioris simétrica no hay mucha variación de los puntos de soporte. Las observaciones anteriores son válidas para las dos distribuciones a prioris bajo estudio, usando para discriminar  $\mathbb{D}_2$ -optimalidad.

Además, como en ambas distribuciones a prioris los diseños óptimos son muy parecidos, se hizo un gráfico ilustrativo, para el caso de la distribución a priori  $\pi_2$ . Allí se muestran tanto los puntos de soporte de las diferentes distribuciones a prioris consideradas, como los pesos asociados. Los pesos son proporcionales al radio de cada uno de los círculos que se muestran en la Figura 4.21. Se observa una clara tendencia de a mayor punto de soporte se tiene un mayor peso en cada uno de los diseños obtenidos. Además, en general ocurre que los puntos de soporte son ó muy similares ó tienen tendencia creciente ó decreciente conforme el parámetro  $\delta$  aumenta en las diferentes distribuciones a prioris ya sea simétricas ó asimétricas a la izquierda ó asimétricas a la derecha, respectivamente.

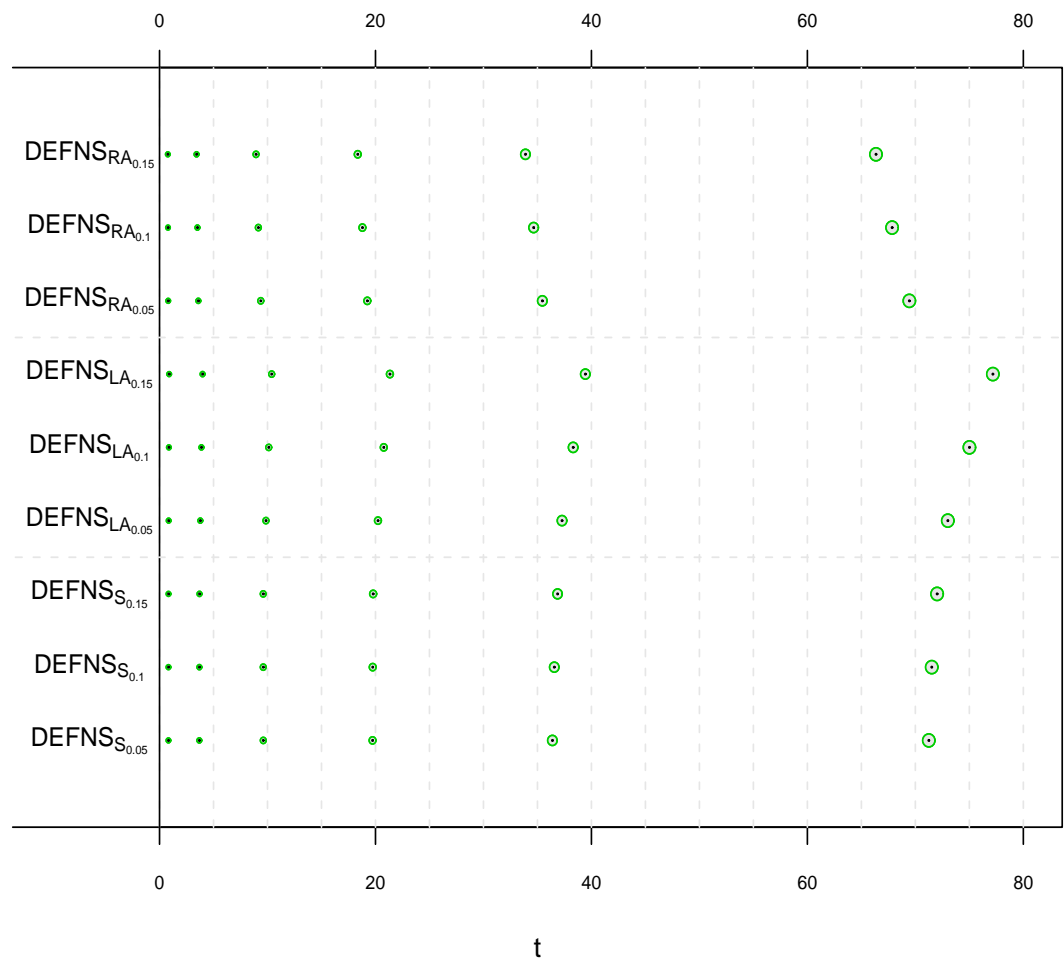


Figura 4.21: Gráfico conjunto de los puntos de soporte, ( $t$ ) y de los pesos asociados a cada uno de los diseños  $DEFNS_{\alpha}$ -óptimos, -proporcional al radio del círculo-, a priori  $\pi_2$ .

# Capítulo 5

## Consideraciones Adicionales

### 5.1. Introducción

En este capítulo se presentan algunos tópicos relacionados con diseños óptimos exactos tanto para el caso de observaciones independientes como de observaciones dependientes. Se verá un método de redondeo para convertir en diseños exactos, los diseños aproximados vistos en el capítulo anterior. A partir del cálculo de las eficiencias se compararán los diseños obtenidos por redondeo con aquellos diseños óptimos exactos en ambos casos, bajo independencia o bajo alguna estructura de correlación en los errores.

Este es un capítulo en desarrollo, donde las ideas de la sección 5.4 fueron inicialmente investigadas en una estancia predoctoral en la Universidad Castilla La Mancha, Ciudad Real, España, junto con los investigadores Jesús López Fidalgo y Mariano Amo Rodríguez. Allí se propone una generalización *ad hoc* del criterio  $\mathbb{T}$ -optimalidad para el caso de observaciones correlacionadas y se combina con los otros criterios en su versión correlacionada.

### 5.2. Implicaciones prácticas

Los diseños obtenidos en éste y los capítulos anteriores fueron diseños óptimos aproximados. Es decir, se tienen las localizaciones de los puntos de experimentación ( $t_i$ ) y la frecuencia  $w_i \in (0, 1)$  de toma de las observaciones en cada punto. Claramente, si se tienen recursos para obtener una muestra de tamaño  $n$ , entonces  $w_i n$ , en general no es un entero. Se necesita algún método de redondeo que garantice que el nuevo diseño (diseño exacto) va a seguir conservando, en la medida de lo posible, las propiedades óptimas del diseño aproximado. Existen procedimientos de redondeo eficientes, por



ejemplo en Pukelsheim (1993) capítulo 12 se presenta el método que se expone a continuación.

Sea un diseño continuo  $\xi$  con pesos  $w_1, w_2, \dots, w_s$ , el método para aproximar este diseño en forma eficiente tiene las siguientes dos fases:

1. Se usa el factor  $n - \frac{1}{2}s$  para calcular  $n_i$  a partir de  $n_i = \lceil (n - \frac{1}{2}s) w_i \rceil$ , donde  $\lceil z \rceil$  denota el entero más pequeño mayor o igual a  $z$ .
2. Se itera hasta que la discrepancia  $d_i = (\sum_{i \leq s} n_i) - n$  sea cero, ya sea incrementando  $n_j$  a  $n_j + 1$  si  $n_j/w_j = \min_{i \leq s} n_i/w_i$  o reduciendo  $n_k$  en una unidad si  $(n_k - 1)/w_k = \max_{i \leq s} (n_i - s)/w_i$ .

Aplicando el método anterior se muestra en Pukelsheim (1993), que el diseño obtenido es en algún sentido eficiente. Se observa que el procedimiento anterior no tiene en cuenta el criterio de optimalidad subyacente, sin embargo el método propuesto funciona bien.

Al aplicar el método anterior a los diseños óptimos locales encontrados para discriminar y estimar simultáneamente se obtienen los diseños dados en la Tabla 5.1 para diferentes tamaños de muestra. En ambos casos, cuando  $n = 6$ , los diseños exactos no tienen réplicas en cada uno de los puntos de soporte. Para los demás valores de  $n$ , se requieren réplicas en diferentes tiempos, dependiendo del valor de la magnitud de  $w_i$ .

En general los diseños obtenidos para el ejemplo en estudio, se pueden interpretar en el siguiente contexto. Considere el diseño exacto  $\xi = \begin{bmatrix} t_1 & \dots & t_6 \\ n_1/n & \dots & n_6/n \end{bmatrix}$ . Además, suponga que se tiene una población de individuos homogéneos con respecto a una característica de interés, es decir, comparten medidas de determinadas covariables muy similares. Se selecciona una muestra de  $n$ -individuos de esta población. En el tiempo  $t = 0$ , a los  $n$ -individuos se les administra el medicamento de estudio. Luego en el tiempo  $t_i$  se toman las observaciones a  $n_i$  individuos de la muestra. Con este procedimiento se tomaría una sola observación de cada individuo. Una de las dificultades del procedimiento es garantizar la homogeneidad de los individuos de la población, que en situaciones prácticas es muy restrictivo.

Otra forma de abordar el problema anterior y así evitar tanto el supuesto de independencia de los errores como las repeticiones, réplicas, en cada uno de los individuos, es considerar la farmacocinética individual, donde se cuenta con un solo individuo a quien se le ha administrado el medicamento. Se desean determinar  $n$  tiempos de muestreo,  $\tau_n = [t_1, \dots, t_n]$ , de aquí en adelante será llamado un diseño exacto, sin réplicas en cada uno de los tiempos, tal que se pueda con estas medidas

	Diseño óptimo $DEFNS_{0.5}^D$	Diseño óptimo $DEFNS_{0.5}^T$
$n$	$\begin{bmatrix} 0.82 & 3.68 & 9.61 & 19.72 & 36.33 & 71.16 \\ 0.10 & 0.11 & 0.13 & 0.15 & 0.22 & 0.29 \end{bmatrix}$	$\begin{bmatrix} 0.87 & 3.71 & 10.09 & 22.86 & 35.90 & 62.94 \\ 0.09 & 0.09 & 0.13 & 0.22 & 0.04 & 0.43 \end{bmatrix}$
6	1 1 1 1 1 1	1 1 1 1 1 1
8	1 1 1 1 2 2	1 1 1 2 1 2
10	1 1 1 2 2 3	1 1 1 2 1 4
20	2 2 3 3 4 6	2 2 3 4 1 8
40	4 5 5 6 9 11	4 4 5 9 2 16
80	9 9 11 12 17 22	7 8 10 18 3 34

Tabla 5.1: Distribución del tamaño de muestra  $n$  eficiente para dos diseños óptimos locales.

discriminar y estimar simultáneamente las funciones no lineales de interés. Lo anterior se verá en la siguiente sección donde se hallan los diseños  $\mathbb{D}$ -óptimos exactos y los diseños  $DEFNS_{\alpha}$ -óptimos para discriminar y estimar en el caso bajo consideración.

### 5.3. Diseños óptimos exactos, bajo independencia

En este caso, el propósito es hallar diseños óptimos exactos, sin replicación, denotados por:  $\tau_n = [t_1 \dots t_n]$ , donde  $t_i \neq t_j$ , para  $i \neq j$ ; tal que maximicen algún funcional de la matriz de información  $M(\tau_n) = \frac{1}{n} \sum_{i=1}^n f(t_i) f^T(t_i)$ . Para la estimación óptima de  $\Theta$  en el modelo  $I$ , se necesitan hallar diseños óptimos exactos que maximicen:

$$|M(\tau_n)|,$$

en este caso se conoce de la teoría de diseños  $\mathbb{D}$ -óptimos que con  $n = m$ ,  $m$ : número de parámetros del modelo, se obtienen diseños óptimos exactos sin réplicas, y coinciden con los diseños  $\mathbb{D}$ -óptimos obtenidos usando diseños aproximados.

En general, con los otros criterios no sucede lo mismo, por ejemplo con  $\mathbb{L}$ -optimalidad, y por ende en el criterio propuesto,  $DEFNS_{\alpha}$ -optimalidad. Con el fin de construir diseños no singulares, es decir, aquellos donde la matriz de información sea invertible, se opta por la construcción de diseños óptimos exactos a partir de seis puntos de soporte, número de parámetros del modelo  $I$ , tomando  $n = 6, 8, 10$ . Con propósitos ilustrativos, se consideró  $\mathbb{T}$ -optimalidad para discriminar y  $\mathbb{L}$ -optimalidad para estimar las funciones no lineales de interés.

Los diseños óptimos obtenidos, tanto exactos como aproximados (redondeados), se hallan en la Tabla 5.2, para tres tamaños de muestra,  $n = 6, 8, 10$ . En la última columna se encuentra la eficiencia de los diseños aproximados con respecto al criterio de optimalidad usado para hallar los diseños exactos. Se observa que los diseños

$n$	Tiempos del diseño óptimo	Eficiencia
6	$\xi_E = [0.84, 3.48, 9.37, 21.84, 55.98, 67.90]$	85.45 %
	$\xi_{Ap} = [0.87, 3.71, 10.09, 22.86, 35.90, 62.94]$	
8	$\xi_E = [0.85, 3.56, 9.85, 22.61, 22.62, 53.07, 66.34, 66.24]$	93.62 %
	$\xi_{Ap} = [0.87, 3.71, 10.09, 22.86, 22.86, 35.90, 62.94, 62.94]$	
10	$\xi_E = [0.85, 3.60, 9.95, 22.04, 22.04, 29.38, 62.29, 62.29, 62.29, 62.29]$	98.99 %
	$\xi_{Ap} = [0.87, 3.71, 10.09, 22.86, 22.86, 35.90, 62.94, 62.94, 62.94, 62.94]$	

Tabla 5.2: Diseños óptimos exactos ( $\xi_E$ ) y aproximados ( $\xi_{Ap}$ ), usando para discriminar  $\mathbb{T}$ -optimalidad y para estimar las funciones no lineales  $\mathbb{L}$ -optimalidad. Eficiencia del diseño aproximado con respecto al diseño exacto,  $\alpha = 0.5$

aproximados, una vez redondeados, tienen eficiencias competitivas frente a los diseños exactos. Además, en cada diseño exacto,  $n = 8, 10$ , se observan tiempos muy parecidos aproximadamente, en los mismos tiempos del diseño aproximado.

## 5.4. Diseños óptimos exactos, con estructura de covarianza

La construcción de los diseños óptimos depende de los supuestos del modelo bajo consideración. Los diseños óptimos de las secciones previas se hallaron usando el supuesto de independencia de los errores. En esta sección se dará una introducción breve de los diseños óptimos cuando el supuesto de independencia no se cumple. Se presentarán algunos diseños que se han construido hasta el momento bajo una estructura de covarianza particular asumida para los errores.

Se considera el modelo de regresión no lineal general:

$$Y(t) = \eta(t, \Theta) + \epsilon(t), \quad t \in \mathcal{X},$$

donde las  $\epsilon(t)$  son variables aleatorias distribuidas normalmente con media cero y covarianza  $Cov(\epsilon(t), \epsilon(t')) = c(t, t')$ ,  $c(\cdot, \cdot)$  es una función conocida. Además,  $\eta(t; \Theta)$  puede ser una de dos funciones parcialmente conocidas  $\eta_1(t; \Theta_1)$  y  $\eta_2(t; \Theta_2)$ , donde  $\Theta_1 \in \Omega_1 \subseteq \mathbb{R}^6$ , y  $\Theta_2 \in \Omega_2 \subseteq \mathbb{R}^4$  son vectores de parámetros desconocidos.

En la literatura de diseños óptimos exactos se han usado las siguientes dos estructuras de covarianza, ver Stehlík et al (2008) y Pepelyshev (2007):

$$Cov(\epsilon(t), \epsilon(t')) = \sigma^2 \exp(-\delta |t - t'|) \quad (5.1)$$

conocida con el nombre de función de covarianza exponencial o:

$$\text{Cov}(\epsilon(t), \epsilon(t')) = \begin{cases} \sigma^2 \left(1 - \frac{|t-t'|}{\rho}\right) & \text{si } |t-t'| \leq \rho, \\ 0 & \text{en otro caso,} \end{cases} \quad (5.2)$$

conocida con el nombre de función de covarianza triangular. Ambas matrices de covarianza dependen de  $\rho$ , que en este trabajo se supone conocido, aunque se puede considerar dentro de la matriz de información en caso de que también se necesite estimar. En lo que sigue la matriz de covarianza se denota por  $\Sigma$ .

A continuación, el propósito está en encontrar tiempos de muestreo óptimo para un sólo individuo, sin réplicas. Es decir, obtener diseños óptimos exactos,  $\tau_n = [t_1, t_2, \dots, t_n]$  de  $n$ -tiempos diferentes, donde  $n = m$  es el número de parámetros del modelo. Inicialmente, el interés está en estimar el vector de parámetros  $\Theta$  usando el criterio  $\mathbb{D}$ -optimalidad. Esto es, obtener un diseño  $\tau_n = [t_1, t_2, \dots, t_n]$  que maximice:

$$\det(I(\tau_n; \Theta)) := \det(X(\tau_n)^T \Sigma(\tau_n)^{-1} X(\tau_n)), \quad (5.3)$$

donde  $X(\tau_n) = \left[ \frac{\partial n(t_i; \Theta)}{\partial \theta_j} \right]$ , e  $I(\tau_n; \Theta)$  es la matriz de información de Fisher en el caso correlacionado.

Además se usará, ya sea  $\mathbb{L}$ - o  $\mathbb{D}_K$ -optimalidad para hallar diseños óptimos que estimen las funciones no lineales. Es decir, construir un diseño  $\tau_n = [t_1, t_2, \dots, t_n]$  que maximice:  $[Tr(K^T I^{-1}(\tau_n; \Theta) K)]^{-1}$  ( $\mathbb{L}$ -optimalidad) ó maximice  $\det(K^T I^{-1}(\tau_n; \Theta) K)$  ( $\mathbb{D}_K$ -optimalidad). Las columnas de la matriz  $K$  son los gradientes de las funciones no lineales que se han considerado.

Para la construcción de los diseños óptimos es razonable el uso de la estructura de covarianza exponencial y para la selección del valor de  $\rho$  se hicieron diversas pruebas con diseños  $\mathbb{D}$ -,  $\mathbb{L}$ -,  $\mathbb{D}_K$ -óptimos exactos y también diseños combinados para estimar los parámetros y las funciones no lineales:  $\mathbb{DL}$  y  $\mathbb{DD}_K$ ; usando diferentes valores de  $\rho$ ,  $\rho \in \{0.001, 0.01, 0.1, 1, 10\}$ , desde muy alta correlación (valores pequeños de  $\rho$ ) a correlaciones bajas (valores grandes de  $\rho$ ). En la mayoría de los casos se observó que los diseños óptimos encontrados no variaban cuando  $\rho > 10$ , y en otros a partir de  $\rho > 1$ . Además, se construyeron las tablas de eficiencias de los diseños encontrados con diferentes valores de  $\rho$ . En el caso  $\rho = 0.01$ , la mayoría de las eficiencias halladas en todos los diseños fue al menos del 50 % comparado con los otros valores de  $\rho$ . También se observó que el criterio  $\mathbb{D}_K$ -optimalidad combinado con  $\mathbb{D}$ -optimalidad produjo, para este valor de  $\rho$ , eficiencias superiores al 90 %, en los dos modelos considerados.

Lo anterior, justifica la elección de  $\rho = 0.01$  para la construcción de los diseños óptimos exactos, es decir, donde la estructura de correlación para las observaciones es

relativamente alta. En la Tabla 5.3 se muestran los diseños óptimos exactos para la estimación de los parámetros y de las funciones no lineales en cada uno de los modelos. Se observa que en el modelo  $I$ ,  $MI$ , los diseños  $\mathbb{D}$  y  $\mathbb{D}_K$  coinciden, debido a que la matriz  $K$  asociada es cuadrada de orden 4, de donde por propiedades del determinante, el criterio a maximizar es el mismo que para  $\mathbb{D}$ -optimalidad. No sucede lo mismo para el modelo  $II$ ,  $MII$ , donde la matriz  $K$  es de orden  $4 \times 6$ .

$\mathbb{D}$ -optimalidad		$\mathbb{D}_K$ -optimalidad	
Modelo	Diseño óptimo exacto: $\tau_m$	Modelo	Diseño óptimo exacto: $\tau_m$
$MII$	[1.22, 3.35, 8.65, 24.16]	$MII$	[1.22, 3.35, 8.65, 24.16]
$MI$	[1.17, 3.20, 7.65, 15.73, 30.62, 57.25]	$MI$	[1.17, 3.15, 7.47, 16.74, 34.33, 55.99]

Tabla 5.3: Diseños óptimos exactos para estimar los parámetros y las funciones no lineales por separado, con  $\rho = 0.01$ .

Para el caso de los diseños combinados, el criterio a maximizar es:

$$\phi_\alpha(\tau_n) = \alpha * \log \phi_1(\tau_n) + (1 - \alpha) \log \phi_2(\tau_n),$$

donde  $\phi_1$  y  $\phi_2$  son los diseños cóncavos que se van a combinar. En la Tabla 5.4 se tienen los diseños óptimos combinados exactos, con  $\alpha = 0.5$ , y  $\rho = 0.01$ , donde  $\phi_1(\tau_n)$  es el criterio  $\mathbb{D}$ -optimalidad y  $\phi_2(\tau_n)$  el criterio  $\mathbb{D}_K$ - o  $\mathbb{L}$ -optimalidad, según el caso.

$\mathbb{D}\mathbb{D}_K$ -optimalidad		$\mathbb{D}\mathbb{L}$ -optimalidad	
Modelo	Diseño óptimo exacto: $\tau_m$	Modelo	Diseño óptimo exacto: $\tau_m$
$MII$	[1.22, 3.35, 8.65, 24.16]	$MII$	[0.53, 2.12, 6.31, 19.08]
$MI$	[1.17, 3.17, 7.58, 16.09, 31.22, 56.41]	$MI$	[0.50, 1.97, 5.65, 12.16, 25.73, 48.45]

Tabla 5.4: Diseños óptimos exactos combinados, usando  $\alpha = 0.5$ , para estimar los parámetros y las funciones no lineales, con  $\rho = 0.01$ .

También se construyeron diseños  $\mathbb{T}$ -óptimos exactos a partir de la maximización del criterio  $\phi_T(\tau_n; \rho)$ , bajo el supuesto de que el modelo  $I$  es el correcto, y  $\Theta_1$  es dado.

$$\begin{aligned} \phi_T(\tau_n; \rho) &= \min_{\Theta_2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^- [\eta_1(t_i; \Theta_1) - \eta_2(t_i; \Theta_2)] [\eta_1(t_j; \Theta_1) - \eta_2(t_j; \Theta_2)] \\ &= \min_{\Theta_2} [\eta_1(\tau_n; \Theta_1) - \eta_2(\tau_n; \Theta_2)]^T \Sigma^{-1}(\tau_n) [\eta_1(\tau_n; \Theta_1) - \eta_2(\tau_n; \Theta_2)] \quad (5.4) \end{aligned}$$

donde  $\sigma_{ij}^-$  es la componente  $(i, j)$  de la matriz  $\Sigma^{-1}$ . El criterio anterior resulta como una generalización *ad hoc* del criterio  $\mathbb{T}$ -optimalidad para el caso correlacionado. Se

halló numéricamente los tiempos de muestreo óptimos, con  $\rho = 0.01$ , dado por: 1.91, 3.75, 8.36, 18.20, 42.904, 110.18.

Finalmente, con los criterios  $\mathbb{T}$ -,  $\mathbb{D}$ -,  $\mathbb{D}_K$ - y  $\mathbb{L}$ -optimalidad en su versión correlacionada, se puede reescribir el criterio de optimalidad propuesto bajo el supuesto de observaciones independientes. Basta con reemplazar la matriz de información de Fisher por su respectiva matriz en función de la matriz de varianzas covarianzas de los errores,  $\Sigma$ . Es decir,

$$\begin{aligned} \phi_\alpha(\tau_n) = & \alpha \log \phi_T(\tau_n; \rho) + \frac{1-\alpha}{4} \left\{ \frac{1}{6} \log |I_1(\tau_n; \Theta_1)| + \frac{1}{4} \log |I_2(\tau_n; \Theta_2)| \right\} \\ & - \frac{1-\alpha}{4} \left\{ \log \text{Tr} [K_1^T I_1^{-1}(\tau_n; \Theta_1) K_1] + \log \text{Tr} [K_2^T I_2^{-1}(\tau_n; \Theta_2) K_2] \right\} \end{aligned}$$

donde  $I_i^{-1}(\tau_n; \Theta_i) = (X_i(\tau_n)^T \Sigma(\tau_n)^{-1} X_i(\tau_n))^{-1}$ , representa la matriz de Fisher del  $i$ -ésimo modelo cuando hay presente estructura de covarianza en los errores.

Se está suponiendo estructura de covarianza exponencial, con  $\rho = 0.01$ , fijo;  $\alpha = 0.5$  y los mismos valores locales de los parámetros que se han estado usando. En la Tabla siguiente se hallan los diseños exactos obtenidos cuando se combina  $\mathbb{T}$ -optimalidad con  $\mathbb{L}$ - ó con  $\mathbb{D}_{K_i}$ -optimalidad para la estimación de las funciones no lineales, en su versión correlacionada, igualmente tomando  $\rho = 0.01$ . Ver Tabla 5.5. Se presentan eficiencias relativamente altas del diseño óptimo bajo el supuesto de observaciones independientes con respecto al criterio de optimalidad en el caso de observaciones correlacionadas con estructura de covarianza exponencial, al menos del 84 % en ambos casos,  $\mathbb{T}\mathbb{L}$  o  $\mathbb{T}\mathbb{D}$ -optimalidad.

Criterio de optimalidad	Diseño óptimo	Eficiencia
$\mathbb{T}\mathbb{L}$ -optimalidad	[1.58, 3.61, 8.27, 18.19, 36.33, 50.43]	84.34 %
$\mathbb{T}\mathbb{D}$ -optimalidad	[1.51, 3.64, 8.34, 18.20, 42.01, 82.17]	89.72 %

Tabla 5.5: Diseños óptimos exactos combinados, usando  $\alpha = 0.5$ , para discriminar, estimar los parámetros y las funciones no lineales, con  $\rho = 0.01$ . Cálculo de la eficiencia del diseño aproximado respectivo.

# Capítulo 6

## Conclusiones

Los diseños óptimos son una buena alternativa para la búsqueda de puntos de toma de muestra en diferentes áreas. Inicialmente se hizo una introducción a la teoría de los diseños óptimos, donde se presentó una versión unificada de los teoremas de equivalencia. También se ampliaron varios ejemplos donde se construyeron diseños óptimos para varios modelos lineales y no lineales, con los respectivos teoremas de equivalencia se verificó que los diseños obtenidos en efecto eran los diseños óptimos.

En esta investigación se construyó un criterio de optimalidad general, que tiene en cuenta múltiples objetivos tales como la discriminación entre dos modelos no lineales, la estimación de los parámetros en cada uno de los modelos y la estimación de funciones no lineales. El primer análisis que se hizo fue bajo el supuesto de independencia de las observaciones. En este caso se consideraron diseños óptimos aproximados donde fue posible enunciar y demostrar el teorema de equivalencia asociado al criterio propuesto. Este resultado fue usado para verificar que los diseños encontrados en efecto eran óptimos. El análisis del criterio se hizo a partir de consideraciones particulares con relación a la ponderación dada a los criterios de optimalidad individuales, aun así se pueden considerar otras ponderaciones, y formular y demostrar el respectivo teorema de equivalencia; por ejemplo considerar como ponderación para la estimación de las funciones no lineales el doble de la ponderación usada para la estimación del vector de parámetros. En el ejemplo de ilustración se consideraron dos modelos anidados, sin embargo, el criterio propuesto es válido en el caso general de modelos no anidados, en su momento se hizo la respectiva observación.

En el caso de estudio, para los modelos resultantes de dos modelos compartimentales, uno de cuatro y el otro de tres compartimentos, se obtuvieron cuatro variantes del criterio propuesto. Se encontraron diseños óptimos locales para cada uno de los cuatro criterios, fijando  $\alpha = 0.50$ , es decir, una ponderación del 50% para el procedimiento de discriminación. Se concluye, que en general, no es posible escoger un criterio de

optimalidad entre los cuatro que cumpla con propiedades de robustez con respecto a la eficiencia de los diseños, es decir, que el diseño óptimo construido por el respectivo criterio tenga eficiencias altas en todos los otros criterios considerados, incluyendo los criterios individuales de discriminación entre los dos modelos y estimación de los parámetros. Cada uno de los criterios considerados tiene buenas eficiencias para discriminación cuando se compara con el respectivo criterio de discriminación individual, además de eficiencias aceptables para la estimación de los parámetros. Se propuso una metodología para encontrar diseños híbridos a partir de los diseños óptimos obtenidos por cada uno de los cuatro criterios y se logró construir diseños con más puntos de soporte y con eficiencias competitivas con respecto a los demás criterios considerados, los cuatro criterios combinados y los criterios individuales para estimación y discriminación entre los dos modelos.

Se consideraron varias distribuciones a prioris, definidas de tal forma que permitieran analizar el efecto de la distribución a priori sobre los diseños óptimos encontrados. Mas aun, fueron usadas con el fin de analizar los cambios los diseños óptimos frente a un posible cambio en las magnitudes de los valores locales del vector de parámetros. Se concluye, para el ejemplo en estudio, que los diseños que usan  $\mathbb{D}_2$ -optimalidad para discriminar y  $\mathbb{L}$ -optimalidad para estimar las funciones no lineales, son muy parecidos a los diseños óptimos obtenidos con las distribuciones a prioris simétricas estudiadas; es decir, para este criterio, los diseños óptimos obtenidos son robustos al valor local usado, permitiendo hasta un margen de error del  $\delta = 15\%$  en la especificación del vector de parámetros. Esto no ocurre en el caso de  $\mathbb{T}$ -optimalidad, donde el quinto punto de soporte tiende a tener pesos pequeños comparado con los pesos de los otros puntos de soporte.

También se analizó la eficiencia de los diseños aproximados convertidos a diseños exactos a partir de un procedimiento de redondeo. Se mostró que estos diseños poseen eficiencias competitivas, al menos del  $85\%$  con respecto a los diseños exactos en el caso de observaciones independientes, con diferentes tamaños de muestra.

Se hizo un análisis inicial de lo que ocurre con los diseños óptimos en el escenario de observaciones dependientes, con una estructura de covarianza exponencial que depende de un parámetro  $\rho$ . Se hallaron los diseños óptimos con un valor fijo de  $\rho$  en 0.01 y se mostró que la eficiencia de los diseños aproximados en el caso independiente, tienen una eficiencia para el caso correlacionado, de al menos un  $84\%$ . Esto se hizo para los casos donde se usó para discriminar  $\mathbb{T}$ -optimalidad y para estimación de las funciones no lineales ya sea  $\mathbb{L}$ - ó  $\mathbb{D}_K$ -optimalidad.

Los diseños óptimos aproximados se hallaron numéricamente a partir de algoritmos



propuestos en la literatura para la obtención de estos ( Fedorov y Hackl (1997)); adicionalmente se propuso un método alternativo para hallar los diseños óptimos, que se ha mostrado, al menos en los diseños hallados en esta tesis, ser más rápido que los hallados usando los algoritmos tradicionales. Con los ejemplos, para diseños aproximados se mostró que el procedimiento alternativo funciona bien, ya que todos los diseños obtenidos verificaron el teorema de equivalencia respectivo. Este mismo procedimiento alternativo, fue usado para obtener los diferentes diseños exactos, y a partir de diferentes diseños iniciales se verificó que en efecto el diseño óptimo obtenido era el mismo.

Líneas de investigación:

Diseños óptimos en farmacocinética poblacional univariada. Se necesita hacer investigación en el caso donde los parámetros del modelo, por ejemplo, las tasas de transferencia no son constantes sino que varían de individuo a individuo. Es decir, hacer un análisis de diseños óptimos, con los mismos objetivos: discriminación y estimación de funciones no lineales, para modelos no lineales mixtos donde el vector de parámetros tiene una distribución conocida. Para el caso de una respuesta univariada y estimación del vector promedio de los parámetros, ver los trabajos de Stroud et al. (2001).

Diseños óptimos en farmacocinética poblacional multivariada. Se necesita hacer más investigación en el caso de observaciones correlacionadas, por ejemplo generalizar el criterio de optimalidad para el caso donde se tienen  $d$ -respuestas en cada tiempo para cada individuo, en el caso del ejemplo estudiado, si además de tener la medida de la concentración en el compartimiento central se tiene la medida de un compartimiento periférico. En Gueorguieva, et al. (2007) y Gueorguieva, et al. (2006) estudian diseños óptimos para el caso donde se tienen  $d$ -respuestas para un mismo individuo y los parámetros son considerados aleatorios.

Diseños óptimos con diferentes estructuras de covarianza. Se propone hacer un estudio de los diseños óptimos con observaciones correlacionadas pero donde en la matriz de información también se considere el parámetro  $\rho$  de la matriz de covarianza. Hacer un estudio comparativo con diferentes estructuras de covarianza, con el fin de responder a la pregunta ¿qué efectos tienen diferentes estructuras de covarianza en los diseños óptimos obtenidos?.

Diseños óptimos Bayesianos. Estudiar los diseños óptimos para múltiples objetivos a partir de la construcción de una función de utilidad apropiada que reúna los diferentes objetivos. Para el caso lineal, Spezzaferri (1988) propuso una función de utilidad para discriminar entre dos modelos y estimar los parámetros, usó función de utilidad

cuadrática modificada.

Diseños óptimos usando medidas de no linealidad. Estudiar la forma como se ven afectados los diseños óptimos cuando se consideran medidas de no linealidad de la superficie esperada, como por ejemplo las medidas de curvatura asociada a los efectos de los parámetros y la medida de curvatura intrínseca. Trabajos en esta dirección son propuestos por: O'Brien y Rawlings (1996) y más recientemente por Diamon y Goto (2007).

# Apéndice A

## A.1. Algoritmo para $\phi$ -optimalidad

Para  $\phi$  un criterio de optimalidad cóncavo y diferenciable, según Sección 3.4; con función de sensibilidad asociada  $D(t; \xi)$ . El siguiente es un algoritmo que considera las diferentes consideraciones propuestas en Fedorov y Hackl (1997) pág. 45 – 51 para criterios de optimalidad cóncavos.

Se inicia con un diseño inicial  $\xi_0$ , tal que  $\phi(\xi_0) < \infty$ ,

1. Hallar  $t_0^+$  tal que:

$$t_0^+ = \arg \max_{t>0} D(t; \xi_0) = \arg \min_{t>0} \{-D(t; \xi_0)\},$$

2. hallar  $\alpha'_0 = \arg \max_{\alpha} \Delta_2 \left( (1 - \alpha) \xi_0 + \alpha \xi_{t_0^+} \right)$ , siendo  $\xi_{t_0^+}$  una medida concentrada en el punto  $t_0^+$ .
3. encontrar  $t_0^-$  y  $t_0$  definidos como:

$$t_0^- = \arg \max_{t \in \text{Supp}(\xi_0)} \{-D(t; \xi_0)\}, \quad t_0 = \arg \min [-D(t_0^+; \xi_0), D(t_0^-; \xi_0)],$$

4. construir el nuevo diseño:

$$\xi_1 = (1 - \alpha_0) \xi_0 + \alpha_0 \xi_{t_0},$$

con:

$$\alpha_0 = \begin{cases} \alpha'_0, & \text{si } t_0 = t_0^+ \\ -\min \left( \alpha'_0, \frac{\xi_0(t_0^-)}{1 - \xi_0(t_0^-)} \right), & \text{si } t_0 = t_0^-. \end{cases}$$

5. hacer  $\xi_0 = \xi_1$ ,
6. ir al paso 1 y continuar hasta que algún criterio de convergencia se cumpla.

Los  $\alpha_s$ ' iniciales se toman de tal forma que satisfagan la siguiente relación:

$$\alpha'_s = \arg \min_{\alpha} -\phi [(1 - \alpha)\xi_s + \alpha\xi(x_s)].$$

En la siguiente sección se presenta el algoritmo para el caso del diseño  $\mathbb{T}$ -óptimo. Difiere de los demás criterios al no involucrar el cálculo de la matriz de información.

## A.2. Algoritmo para $\mathbb{T}$ -optimalidad

Dado un diseño inicial  $\xi_0$ ,  $\Theta_1^0$  conocido. Considere las funciones  $\Psi_2(t; \xi_0)$  y  $\Delta_2(\xi_0)$  definidas en la sección 3.7, página 45.

1. Hallar  $\Theta_2^* = \arg \min_{\Theta_2} \int_{\mathcal{X}} \{\eta_{1T}(\Theta_1^0; t) - \eta_2(\Theta_2; t)\}^2 d\xi_0(t)$ ,
2. hallar  $t_0^+$  tal que:

$$t_0^+ = \arg \max_{t>0} \Psi_2(t; \xi_0) = \arg \min_{t>0} \{-\Psi_2(t; \xi_0)\},$$

3. hallar  $\alpha'_0 = \arg \max_{\alpha} \Delta_2((1 - \alpha)\xi_0 + \alpha\xi_{t_0^+})$ ,
4. encontrar  $t_0^-$  y  $t_0$  definidos como:

$$t_0^- = \arg \max_{t \in \text{Supp}(\xi_0)} \{-\Psi_2(t, \xi_0)\}, \quad t_0 = \arg \min [-\Psi_2(t_0^+, \xi_0), \Psi_2(t_0^-, \xi_0)],$$

5. construir el nuevo diseño:

$$\xi_1 = (1 - \alpha_0)\xi_0 + \alpha_0\xi_{t_0},$$

con:

$$\alpha_0 = \begin{cases} \alpha'_0, & \text{si } t_0 = t_0^+ \\ -\min\left(\alpha'_0, \frac{\xi_0(t_0^-)}{1 - \xi_0(t_0^-)}\right), & \text{si } t_0 = t_0^-. \end{cases}$$

6. hacer  $\xi_0 = \xi_1$ , y continuar sucesivamente hasta que algún criterio de convergencia sea satisfecho.

## A.3. Procedimiento alternativo

A continuación se expone una forma genérica de las funciones consideradas para optimización. En general lo que busca el procedimiento es optimizar una función en  $2s$  parámetros, e internamente estandarizar los últimos  $s$ -valores con el fin de que sumen la unidad. Al final del algoritmo se presenta un ejemplo programado en el lenguaje  $R$ .

1. Sea  $s$  el número de puntos de soporte del diseño óptimo,  $s$  es conocido.
2. Defina inicialmente la función de optimalidad donde uno de sus argumentos es el diseño,  $\xi$ . Denote esta función por  $\phi$ .
3. Defina una nueva función, cuyo argumento sea un vector,  $v$ , de dimensión  $2 \times s$ . Allí estandarice las últimas  $s$ -componentes de  $v$ , calcule el diseño  $\xi$  y luego evalúe la función  $\phi$ .
4. Por último use algún optimizador para minimizar la función anterior.

En el siguiente ejemplo se muestra como funciona para  $\mathbb{D}$ -optimalidad, en el caso del modelo cuadrático, donde  $x \in [0, 1]$ .

#1. Se define la función  $f(x)$ :

```
fx <- function(x) as.matrix(c(1,x,x^2))
```

#2. Cálculo de la matriz de momentos,  $\xi$  es una matriz de orden  $2 \times s$

```
Mom <- function(xi)
```

```
{
```

```
  x <- as.matrix(xi[1,]);  w <- xi[2,]
```

```
  X <- t(apply(x,1,fx));  W <- diag(w); Mom <- t(X)%*%W%*%X
```

```
  return(Mom)
```

```
}
```

#3. Definición de la función a minimizar:  $-\log(|M(\xi)|)$

```
phi <- function(xi) -log(det(Mom(xi)))
```

#4. La función apropiada a minimizar: con argumento un vector de

```
# dimension 2s, las primeras s componentes son los puntos de
```

```
# soporte y los últimos s son los pesos.
```

```
g <- function(dis)
```

```
{
```

```
  ss <- length(dis); s <- ss/2
```

```
  xi <- matrix( c( dis[1:s],dis[(s+1):ss]/sum(dis[(s+1):ss]  )),ncol=s,byrow=T)
```

```
  au <- phi(xi)
```

```
  return(au)
```

```
}
```

#5. Se minimiza  $g$ :

```
#Valores iniciales del diseño:
```

```
disini<-c(0.2,0.8,0.9,0.2,0.3,0.5)
```

```
ss <-length(disini)
```

```
s <-ss/2
```

```
res <-nlminb(disini, g, lower =c(rep(0,ss)), upper = c(rep(1,ss)))
```

```
soportoxicand <- res$par[1:s] # 0.0 0.5 1.0  
pesosxicand <- res$par[(s+1):ss]/sum(res$par[(s+1):ss]) # 0.33 0.33 0.33
```

# Apéndice B

## B.1. Gradiente de la función $\eta_C(t; \Theta)$ , modelo I

El gradiente de la función  $\eta_C(t; \Theta)$  está dado por:

$$f(t; \Theta)^T = \nabla \eta_C^T(t; \Theta) = [f_1(t; \Theta) \dots f_6(t; \Theta)],$$

donde:

1.  $f_1(t; \Theta) = \frac{\partial \eta_1}{\partial \theta_1} = A_0 \left\{ h_1(\Theta) e^{-\theta_1 t} + \sum_{j=1}^3 h_{j+1}(\Theta) e^{\lambda_j t} \right\} + \theta_1 A_0 Expr_0$ , siendo:

$$Expr_0 = -h_1(\Theta) t e^{-\theta_1 t} + \frac{\partial h_1(\Theta)}{\partial \theta_1} e^{-\theta_1 t} + \sum_{j=1}^3 \left( \frac{\partial h_{j+1}(\Theta)}{\partial \theta_1} e^{\lambda_j t} \right).$$

2.  $f_i(t; \Theta) = \frac{\partial \eta_C}{\partial \theta_i} = \theta_1 A_0 \left\{ \frac{\partial h_1(\Theta)}{\partial \theta_i} e^{-\theta_1 t} + \sum_{j=1}^3 \left[ h_{j+1}(\Theta) t \frac{\partial \lambda_j}{\partial \theta_i} e^{\lambda_j t} + \frac{\partial h_{j+1}(\Theta)}{\partial \theta_i} e^{\lambda_j t} \right] \right\}$ , válido para  $i = 2, \dots, 6$ .

## B.2. Características de interés, modelo I.

En esta sección se presentan las expresiones para el área bajo la curva de concentración, el tiempo para la concentración máxima, la concentración máxima, y el primer tiempo donde la concentración alcanza un porcentaje preestablecido de la concentración máxima. También se dan las derivadas, en el caso donde las observaciones se toman en el compartimiento C.

1.  $H_1(\Theta) = Area = \int_0^\infty \eta_C(t; \Theta) dt = -\frac{\theta_3 \theta_5 A_0}{\lambda_1 \lambda_2 \lambda_3} = -\frac{\theta_3 \theta_5 A_0}{-\theta_3 \theta_5 \theta_6} = \frac{A_0}{\theta_6}$ . Además en este caso el gradiente de  $H_1$  está dado por:

$$\nabla H_1 = [0 \ 0 \ 0 \ 0 \ 0 \ -A_0/\theta_6^2]^T.$$

2.  $H_2(\Theta) = t_{m\acute{a}x}$ , no se puede encontrar una expresión cerrada.  $H_2(\Theta)$  se obtiene resolviendo numéricamente la siguiente ecuación:

$$\frac{\partial \eta_C(t; \Theta)}{\partial t} = 0 = \theta_1 A_0 \left[ -\theta_1 h_1(\Theta) e^{-\theta_1 t} + \lambda_1 h_2(\Theta) e^{\lambda_1 t} + h_3(\Theta) \lambda_2 e^{\lambda_2 t} + h_4(\Theta) \lambda_3 e^{\lambda_3 t} \right].$$

En este caso el gradiente de  $H_2$ ,  $\nabla H_2$ , se halla derivando implícitamente la ecuación anterior y luego despejando  $\frac{\partial H_2(\Theta)}{\partial \theta_i}$ ,  $i = 1, \dots, 6$ . Tomando  $t = t_{\text{máx}}$ , se muestra que:

$$\frac{\partial H_2(\Theta)}{\partial \theta_1} = \frac{-\left(\theta_1 h_1(\Theta) t e^{-\theta_1 t} - \theta_1 \frac{\partial h_1(\Theta)}{\partial \theta_1} e^{-\theta_1 t} - h_1(\Theta) e^{-\theta_1 t} + \sum_{i=1}^3 \lambda_i \frac{\partial h_{i+1}(\Theta)}{\partial \theta_1} e^{\lambda_i t}\right)}{\theta_1^2 h_1(\Theta) e^{-\theta_1 t} + \sum_{i=1}^3 \lambda_i^2 h_{i+1}(\Theta) e^{\lambda_i t}},$$

$$\frac{\partial H_2(\Theta)}{\partial \theta_i} = \frac{\theta_1 \frac{\partial h_1(\Theta)}{\partial \theta_i} e^{-\theta_1 t} - \sum_{j=1}^3 \left[ \lambda_j e^{\lambda_j t} \left( \frac{\partial h_{j+1}(\Theta)}{\partial \theta_i} + h_{j+1}(\Theta) t \frac{\partial \lambda_j}{\partial \theta_i} \right) + h_{j+1}(\Theta) e^{\lambda_j t} \frac{\partial \lambda_j}{\partial \theta_i} \right]}{\theta_1^2 h_1(\Theta) e^{-\theta_1 t} + \sum_{j=1}^3 \lambda_j^2 h_{j+1}(\Theta) e^{\lambda_j t}},$$

para  $i = 2, \dots, 6$ .

Las derivadas de las raíces,  $\lambda_i$ , con respecto a cada uno de los parámetros, están dadas por las siguientes expresiones:

Para abreviar, denote por  $T_1 = \sum_{i \neq 1} \theta_i$ ,  $T_2 = (\theta_2 + \theta_3 + \theta_6)\theta_5 + \theta_3(\theta_4 + \theta_6)$ ;

$$\frac{\partial \lambda_i}{\partial \theta_1} = 0, \quad \frac{\partial \lambda_i}{\partial \theta_2} = \frac{-\lambda_i^2 - \lambda_i \theta_5}{3\lambda_i^2 + 2T_1 \lambda_i + T_2}$$

$$\frac{\partial \lambda_i}{\partial \theta_3} = \frac{-\lambda_i^2 - \lambda_i(\theta_4 + \theta_5 + \theta_6) - \theta_5 \theta_6}{3\lambda_i^2 + 2T_1 \lambda_i + T_2}, \quad \frac{\partial \lambda_i}{\partial \theta_4} = \frac{-\lambda_i^2 - \lambda_i \theta_3}{3\lambda_i^2 + 2T_1 \lambda_i + T_2}$$

$$\frac{\partial \lambda_i}{\partial \theta_5} = \frac{-\lambda_i^2 - \lambda_i(\theta_2 + \theta_3 + \theta_6) - \theta_3 \theta_6}{3\lambda_i^2 + 2T_1 \lambda_i + T_2}, \quad \frac{\partial \lambda_i}{\partial \theta_6} = \frac{-\lambda_i^2 - \lambda_i(\theta_5 + \theta_3) - \theta_3 \theta_5}{3\lambda_i^2 + 2T_1 \lambda_i + T_2}.$$

Las funciones  $h_i(\Theta)$  tienen las siguientes derivadas:

$$\frac{\partial h_2(\Theta)}{\partial \theta_1} = -\frac{(\lambda_1 + \theta_3)(\lambda_1 + \theta_5)}{(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)(\lambda_1 + \theta_1)^2}$$

$$\frac{\partial h_2(\Theta)}{\partial \theta_j} = \frac{(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)(\lambda_1 + \theta_1) \left[ (\lambda_1 + \theta_3) \frac{\partial \lambda_1}{\partial \theta_j} + (\lambda_1 + \theta_5) \frac{\partial \lambda_1}{\partial \theta_j} \right] - Expr1(j)}{(\lambda_3 - \lambda_1)^2 (\lambda_2 - \lambda_1)^2 (\lambda_1 + \theta_1)^2}, j = 2, 4, 6$$

$$Expr1(j) = (\lambda_1 + \theta_3)(\lambda_1 + \theta_5) \left\{ (\lambda_3 - \lambda_1) \left[ (\theta_1 + \lambda_1) \left( \frac{\partial \lambda_2}{\partial \theta_j} - \frac{\partial \lambda_1}{\partial \theta_j} \right) + (\lambda_2 - \lambda_1) \frac{\partial \lambda_1}{\partial \theta_j} \right] \right. \\ \left. + (\lambda_2 - \lambda_1)(\theta_1 + \lambda_1) \left( \frac{\partial \lambda_3}{\partial \theta_j} - \frac{\partial \lambda_1}{\partial \theta_j} \right) \right\} j = 2, \dots, 6$$

$$\frac{\partial h_2(\Theta)}{\partial \theta_3} = \frac{(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)(\lambda_1 + \theta_1) \left[ (\lambda_1 + \theta_3) \frac{\partial \lambda_1}{\partial \theta_3} + (\lambda_1 + \theta_5) \left( \frac{\partial \lambda_1}{\partial \theta_3} + 1 \right) \right] - Expr1(3)}{(\lambda_3 - \lambda_1)^2 (\lambda_2 - \lambda_1)^2 (\lambda_1 + \theta_1)^2}$$

$$\frac{\partial h_2(\Theta)}{\partial \theta_5} = \frac{(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)(\lambda_1 + \theta_1) \left[ (\lambda_1 + \theta_3) \left( \frac{\partial \lambda_1}{\partial \theta_5} + 1 \right) + (\lambda_1 + \theta_5) \frac{\partial \lambda_1}{\partial \theta_5} \right] - Expr1(5)}{(\lambda_3 - \lambda_1)^2 (\lambda_2 - \lambda_1)^2 (\lambda_1 + \theta_1)^2}$$



$$\frac{\partial h_3(\Theta)}{\partial \theta_1} = \frac{(\lambda_2 + \theta_3)(\lambda_2 + \theta_5)}{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_2)(\lambda_2 + \theta_1)^2}$$

$$\frac{\partial h_3(\Theta)}{\partial \theta_j} = \frac{-(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_2 + \theta_1) \left[ \frac{\partial \lambda_2}{\partial \theta_j} (\lambda_2 + \theta_5) + \frac{\partial \lambda_2}{\partial \theta_j} (\lambda_2 + \theta_3) \right] + Expr2(j)}{(\lambda_1 - \lambda_2)^2 (\lambda_2 - \lambda_3)^2 (\lambda_2 + \theta_1)^2}, j = 2, 4, 6,$$

$$Expr2(j) = (\lambda_2 + \theta_3)(\lambda_2 + \theta_5) \left\{ (\lambda_1 - \lambda_2) \left[ \left( \frac{\partial \lambda_2}{\partial \theta_j} - \frac{\partial \lambda_3}{\partial \theta_j} \right) (\lambda_2 + \theta_1) + (\lambda_2 - \lambda_3) \frac{\partial \lambda_2}{\partial \theta_j} \right] \right. \\ \left. + \left( \frac{\partial \lambda_1}{\partial \theta_j} - \frac{\partial \lambda_2}{\partial \theta_j} \right) (\lambda_2 - \lambda_3)(\lambda_2 + \theta_1) \right\}, j = 2, \dots, 6.$$

$$\frac{\partial h_3(\Theta)}{\partial \theta_3} = \frac{-(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_2 + \theta_1) \left[ \left( \frac{\partial \lambda_2}{\partial \theta_3} + 1 \right) (\lambda_2 + \theta_5) + \frac{\partial \lambda_2}{\partial \theta_3} (\lambda_2 + \theta_3) \right] + Expr2(3)}{(\lambda_1 - \lambda_2)^2 (\lambda_2 - \lambda_3)^2 (\lambda_2 + \theta_1)^2}$$

$$\frac{\partial h_3(\Theta)}{\partial \theta_5} = \frac{-(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_2 + \theta_1) \left[ \frac{\partial \lambda_2}{\partial \theta_5} (\lambda_2 + \theta_5) + \left( \frac{\partial \lambda_2}{\partial \theta_5} + 1 \right) (\lambda_2 + \theta_3) \right] + Expr2(5)}{(\lambda_1 - \lambda_2)^2 (\lambda_2 - \lambda_3)^2 (\lambda_2 + \theta_1)^2}$$

$$\frac{\partial h_4(\Theta)}{\partial \theta_1} = -\frac{(\lambda_3 + \theta_3)(\lambda_3 + \theta_5)}{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3)(\lambda_3 + \theta_1)^2}$$

$$\frac{\partial h_4(\Theta)}{\partial \theta_j} = \frac{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3)(\lambda_3 + \theta_1) \left[ \frac{\partial \lambda_3}{\partial \theta_j} (\lambda_3 + \theta_5) + \frac{\partial \lambda_3}{\partial \theta_j} (\lambda_3 + \theta_3) \right] - Expr3(j)}{(\lambda_2 - \lambda_3)^2 (\lambda_1 - \lambda_3)^2 (\lambda_3 + \theta_1)^2}, j = 2, 4, 6,$$

$$Expr3(j) = (\lambda_3 + \theta_3)(\lambda_3 + \theta_5) \left\{ (\lambda_2 - \lambda_3) \left[ \left( \frac{\partial \lambda_1}{\partial \theta_j} - \frac{\partial \lambda_3}{\partial \theta_j} \right) (\lambda_3 + \theta_1) + (\lambda_1 - \lambda_3) \frac{\partial \lambda_3}{\partial \theta_j} \right] \right. \\ \left. + \left( \frac{\partial \lambda_2}{\partial \theta_j} - \frac{\partial \lambda_3}{\partial \theta_j} \right) (\lambda_1 - \lambda_3)(\lambda_3 + \theta_1) \right\}, j = 2, \dots, 6.$$

$$\frac{\partial h_4(\Theta)}{\partial \theta_3} = \frac{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3)(\lambda_3 + \theta_1) \left[ \left( \frac{\partial \lambda_3}{\partial \theta_3} + 1 \right) (\lambda_3 + \theta_5) + \frac{\partial \lambda_3}{\partial \theta_3} (\lambda_3 + \theta_3) \right] - Expr3(3)}{(\lambda_2 - \lambda_3)^2 (\lambda_1 - \lambda_3)^2 (\lambda_3 + \theta_1)^2},$$

$$\frac{\partial h_4(\Theta)}{\partial \theta_5} = \frac{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3)(\lambda_3 + \theta_1) \left[ \frac{\partial \lambda_3}{\partial \theta_5} (\lambda_3 + \theta_5) + \left( \frac{\partial \lambda_3}{\partial \theta_5} + 1 \right) (\lambda_3 + \theta_3) \right] - Expr3(5)}{(\lambda_2 - \lambda_3)^2 (\lambda_1 - \lambda_3)^2 (\lambda_3 + \theta_1)^2}.$$

3.  $H_3(\Theta) = \eta_C(t_{\text{máx}}; \Theta)$ , concentración máxima. El gradiente de esta función está dado por:

$$\nabla H_3 = \left[ \frac{\partial H_3}{\partial \theta_1} \quad \frac{\partial H_3}{\partial \theta_2} \quad \frac{\partial H_3}{\partial \theta_3} \quad \frac{\partial H_3}{\partial \theta_4} \quad \frac{\partial H_3}{\partial \theta_5} \quad \frac{\partial H_3}{\partial \theta_6} \right]^T,$$

donde:

$$\begin{aligned} \frac{\partial H_3(\Theta)}{\partial \theta_1} &= A_0 \left[ h_1(\Theta) e^{-\theta_1 t_{\text{máx}}} + \sum_{j=1}^3 h_{j+1}(\Theta) e^{\lambda_j t_{\text{máx}}} \right] \\ &\quad + \theta_1 A_0 \left[ \frac{\partial (h_1(\Theta) e^{-t_{\text{máx}} \theta_1})}{\partial \theta_1} + \sum_{j=1}^3 \frac{\partial (h_{j+1}(\Theta) e^{t_{\text{máx}} \lambda_j})}{\partial \theta_1} \right] \\ \frac{\partial H_3(\Theta)}{\partial \theta_i} &= A_0 \theta_1 \left[ \frac{\partial (h_1(\Theta) e^{-t_{\text{máx}} \theta_1})}{\partial \theta_i} + \sum_{j=1}^3 \frac{\partial (h_{j+1}(\Theta) e^{t_{\text{máx}} \lambda_j})}{\partial \theta_i} \right], i \neq 1. \end{aligned}$$

$$\frac{\partial (h_1(\Theta) e^{-t_{\text{máx}} \theta_1})}{\partial \theta_1} = h_1(\Theta) e^{-\theta_1 t_{\text{máx}}} \left( -t_{\text{máx}} - \theta_1 \frac{\partial t_{\text{máx}}}{\partial \theta_1} \right) + \frac{\partial h_1(\Theta)}{\partial \theta_1} e^{-t_{\text{máx}} \theta_1}$$

$$\frac{\partial (h_1(\Theta) e^{-t_{\text{máx}} \theta_1})}{\partial \theta_i} = h_1(\Theta) e^{-\theta_1 t_{\text{máx}}} \left( -\theta_1 \frac{\partial t_{\text{máx}}}{\partial \theta_i} \right) + \frac{\partial h_1(\Theta)}{\partial \theta_i} e^{-t_{\text{máx}} \theta_1}, i > 1.$$

$$\begin{aligned} \frac{\partial (h_{j+1}(\Theta) e^{t_{\text{máx}} \lambda_j})}{\partial \theta_i} &= h_{j+1}(\Theta) \left( e^{t_{\text{máx}} \lambda_j} \left( \frac{\partial t_{\text{máx}}}{\partial \theta_i} \lambda_j + t_{\text{máx}} \frac{\partial \lambda_j}{\partial \theta_i} \right) \right) + \frac{\partial h_{j+1}(\Theta)}{\partial \theta_i} e^{t_{\text{máx}} \lambda_j}, \\ &\quad j = 1, 2, 3; \quad i = 1, \dots, 6. \end{aligned}$$

En este caso:  $\frac{\partial t_{\text{máx}}}{\partial \theta_i} = \frac{\partial H_2(\Theta)}{\partial \theta_i}$ ,  $i = 1, \dots, 6$ .

4.  $H_4(\Theta) = T_k$ ,  $0 < k < 1$ , siendo  $T_k$  el primer tiempo donde la concentración en el compartimiento  $C$  alcanza un  $100k\%$  de su concentración máxima. Es decir,  $T_k$  satisface:

$$\eta_C(T_k; \Theta) = k \eta_C(t_{\text{máx}}; \Theta),$$

equivalente a:

$$h_1(\Theta) \left[ e^{-\theta_1 T_k} - k e^{-\theta_1 t_{\text{máx}}} \right] + \sum_{j=1}^3 h_{j+1}(\Theta) \left[ e^{\lambda_j T_k} - k e^{\lambda_j t_{\text{máx}}} \right] = 0. \quad (\text{B.1})$$

Al derivar implícitamente la ecuación B.1 con respecto a cada componente de  $\Theta$ , se obtienen las siguientes expresiones:

$$P_1(\Theta) = -\theta_1 h_1(\Theta) e^{-\theta_1 T_k} + \lambda_1 h_2(\Theta) e^{\lambda_1 T_k} + \lambda_2 h_3(\Theta) e^{\lambda_2 T_k} + \lambda_3 h_4(\Theta) e^{\lambda_3 T_k}$$

$$\begin{aligned} P_2(\Theta) &= \left( T_k h_1(\Theta) - \frac{\partial h_1(\Theta)}{\partial \theta_1} \right) e^{-T_k \theta_1} - \sum_{j=1}^3 \left( \frac{\partial h_{j+1}(\Theta)}{\partial \theta_1} e^{\lambda_j T_k} \right) \\ &\quad + k e^{-\theta_1 t_{\text{máx}}} \left[ \frac{\partial h_1(\Theta)}{\partial \theta_1} - h_1(\Theta) \left( t_{\text{máx}} + \theta_1 \frac{\partial t_{\text{máx}}}{\partial \theta_1} \right) \right] \\ &\quad + \sum_{j=1}^3 k e^{\lambda_j t_{\text{máx}}} \left( \frac{\partial h_{j+1}(\Theta)}{\partial \theta_1} + \lambda_j h_{j+1}(\Theta) \frac{\partial t_{\text{máx}}}{\partial \theta_1} \right) \end{aligned}$$

$$\begin{aligned}
 P_3(\Theta; j) &= ke^{-t_{\max}\theta_1} \left( \frac{\partial h_1(\Theta)}{\partial \theta_j} - \theta_1 h_1(\Theta) \frac{\partial t_{\max}}{\partial \theta_j} \right) \\
 &\quad + k \sum_{i=1}^3 e^{\lambda_i t_{\max}} \left[ \frac{\partial h_{i+1}(\Theta)}{\partial \theta_j} + h_{i+1}(\Theta) \left( t_{\max} \frac{\partial \lambda_i}{\partial \theta_j} + \lambda_i \frac{\partial t_{\max}}{\partial \theta_j} \right) \right], \quad j \geq 2. \\
 P_4(\Theta; j) &= -\frac{\partial h_1(\Theta)}{\partial \theta_j} e^{-\theta_1 T_k} - \sum_{i=1}^3 e^{\lambda_i T_k} \left( \frac{\partial h_{i+1}(\Theta)}{\partial \theta_j} + T_k h_2(\Theta) \frac{\partial \lambda_i}{\partial \theta_j} \right), \quad j \geq 2. \\
 \frac{\partial H_4(\Theta)}{\partial \theta_1} &= \frac{\partial T_k}{\partial \theta_1} = \frac{P_2(\Theta)}{P_1(\Theta)} \\
 \frac{\partial H_4(\Theta)}{\partial \theta_j} &= \frac{P_3(\Theta, j) + P_4(\Theta, j)}{P_1(\Theta)}, \quad j \geq 2.
 \end{aligned}$$

### B.3. Gradiente de la función $\eta_C(t; \mathbb{B})$ , modelo II

El gradiente de la función  $\eta_C(t; \mathbb{B})$  viene dado por:

$$w(t; \mathbb{B})^T = \nabla \eta_C^T(t; \mathbb{B}) = [w_1(\mathbb{B}) \ w_2(\mathbb{B}) \ w_3(\mathbb{B}) \ w_4(\mathbb{B})],$$

donde:

1.  $S(\mathbb{B}) = r_1(\mathbb{B}) e^{-t\beta_1} + r_2(\mathbb{B}) e^{\lambda_1 t} + r_3(\mathbb{B}) e^{\lambda_2 t}$ .
2.  $\frac{\partial r_1(\mathbb{B})}{\partial \beta_1} = -\frac{\partial r_2(\mathbb{B})}{\partial \beta_1} - \frac{\partial r_3(\mathbb{B})}{\partial \beta_1}$ .
3.  $\frac{\partial r_2(\mathbb{B})}{\partial \beta_1} = -\frac{\lambda_1 + \beta_3}{(\lambda_1 + \beta_1)^2}$ .
4.  $\frac{\partial r_3(\mathbb{B})}{\partial \beta_1} = \frac{\lambda_2 + \beta_3}{(\lambda_2 + \beta_1)^2}$ .
5.  $\frac{\partial r_2(\mathbb{B})}{\partial \beta_j} = -\frac{(\lambda_1 + \beta_1) \left( \frac{\partial \lambda_1}{\partial \beta_j} + I_{\{3\}}(j) \right) - (\lambda_2 + \beta_3) \left( \frac{\partial \lambda_2}{\partial \beta_j} \right)}{(\lambda_2 + \beta_1)^2}, \quad j \geq 2$ .
6.  $\frac{\partial r_1(\mathbb{B})}{\partial \beta_j} = -\frac{\partial r_2(\mathbb{B})}{\partial \beta_j} - \frac{\partial r_3(\mathbb{B})}{\partial \beta_j}, \quad j > 1$ .
7.  $\frac{\partial S(\mathbb{B})}{\partial \beta_1} = -tr_1(\mathbb{B}) e^{-t\beta_1} + e^{-t\beta_1} \frac{\partial r_1(\mathbb{B})}{\partial \beta_1} + \frac{\partial r_2(\mathbb{B})}{\partial \beta_1} e^{\lambda_1 t} + \frac{\partial r_3(\mathbb{B})}{\partial \beta_1} e^{\lambda_2 t}$ .
8.  $w_1(\mathbb{B}) = \frac{A_0 \beta_1}{\lambda_1 - \lambda_2} \frac{\partial S(\mathbb{B})}{\partial \beta_1} + S(\mathbb{B}) \frac{A_0}{\lambda_1 - \lambda_2}$ .
9.  $\frac{\partial S(\mathbb{B})}{\partial \beta_j} = e^{-t\beta_1} \frac{\partial r_1(\mathbb{B})}{\partial \beta_j} + \frac{\partial r_2(\mathbb{B})}{\partial \beta_j} e^{\lambda_1 t} + r_2(\mathbb{B}) t \frac{\partial \lambda_1}{\partial \beta_j} e^{\lambda_1 t} + r_3(\mathbb{B}) t \frac{\partial \lambda_2}{\partial \beta_j} e^{\lambda_2 t} + e^{\lambda_2 t} \frac{\partial r_3(\beta)}{\partial \beta_2}, \quad j > 1$ .
10.  $w_j(\mathbb{B}) = \frac{A_0 \beta_1}{\lambda_1 - \lambda_2} \frac{\partial S(\mathbb{B})}{\partial \beta_j} - S(\mathbb{B}) A_0 \beta_1 \frac{\frac{\partial \lambda_1}{\partial \beta_j} - \frac{\partial \lambda_2}{\partial \beta_j}}{(\lambda_1 - \lambda_2)^2}, \quad j = 2, 3, 4$ .

## B.4. Características de interés, modelo II

1.  $G_1(\mathbb{B}) = \text{Área} = \int_0^\infty \eta_C(t; \mathbb{B}) dt = \frac{A_0}{\beta_4}$ . El gradiente de  $G_1$  está dado por:

$$\nabla G_1 = [0 \ 0 \ 0 \ -A_0/\beta_4^2].$$

2.  $G_2(\mathbb{B}) = t_{\text{máx}}$ . Resulta como solución numérica de la ecuación:

$$\frac{\partial \eta_C(t; \mathbb{B})}{\partial t} = 0 = \frac{A_0 \beta_1}{\lambda_1 - \lambda_2} \left[ -\beta_1 r_1(\mathbb{B}) e^{-t\beta_1} + r_2(\mathbb{B}) \lambda_1 e^{\lambda_1 t} + r_3(\mathbb{B}) \lambda_2 e^{\lambda_2 t} \right],$$

equivalente a:

$$\left[ -\beta_1 r_1(\mathbb{B}) e^{-t\beta_1} + r_2(\mathbb{B}) \lambda_1 e^{\lambda_1 t} + r_3(\mathbb{B}) \lambda_2 e^{\lambda_2 t} \right] = 0 \quad (\text{B.2})$$

Y el gradiente de  $G_2(\mathbb{B})$  se halla al derivar implícitamente la ecuación B.2, con respecto a cada parámetro. Se muestra fácilmente que:

$$\begin{aligned} \text{Deno} &= \beta_1^2 r_1(\mathbb{B}) e^{-\beta_1 t_{\text{máx}}} + \lambda_1^2 r_2(\mathbb{B}) e^{\lambda_1 t_{\text{máx}}} + \lambda_2^2 r_3(\mathbb{B}) e^{\lambda_2 t_{\text{máx}}} \\ \text{Num} &= \beta_1 \frac{\partial r_1(\mathbb{B})}{\partial \beta_j} e^{-\beta_1 t_{\text{máx}}} - e^{\lambda_1 t_{\text{máx}}} \left( \lambda_1 \frac{\partial r_2(\mathbb{B})}{\partial \beta_j} + \lambda_1 r_2(\mathbb{B}) t_{\text{máx}} \frac{\partial \lambda_1}{\partial \beta_j} + r_2(\mathbb{B}) \frac{\partial \lambda_1}{\partial \beta_j} \right) \\ &\quad - e^{\lambda_2 t_{\text{máx}}} \left( \lambda_2 \frac{\partial r_3(\mathbb{B})}{\partial \beta_j} + \lambda_2 r_3(\mathbb{B}) t_{\text{máx}} \frac{\partial \lambda_2}{\partial \beta_j} + r_3(\mathbb{B}) \frac{\partial \lambda_2}{\partial \beta_j} \right), \\ \frac{\partial G_2(\mathbb{B})}{\partial \beta_1} &= \frac{\beta_1 e^{-\beta_1 t_{\text{máx}}} \left( \frac{\partial r_1(\mathbb{B})}{\partial \beta_1} - t_{\text{máx}} r_1(\mathbb{B}) + r_1(\mathbb{B}) / \beta_1 \right) - \lambda_1 \frac{\partial r_2(\mathbb{B})}{\partial \beta_1} e^{\lambda_1 t_{\text{máx}}} - \lambda_2 \frac{\partial r_3(\mathbb{B})}{\partial \beta_1} e^{\lambda_2 t_{\text{máx}}}}{\text{Deno}}. \\ \frac{\partial G_2(\mathbb{B})}{\partial \beta_j} &= \frac{\text{Num}}{\text{Deno}}, \quad j \geq 2. \end{aligned}$$

3.  $G_3(\mathbb{B}) = \eta_C(t_{\text{máx}}; \mathbb{B})$ , concentración máxima. Su gradiente asociado está dado por:

$$\begin{aligned} S(\mathbb{B}) &= r_1(\mathbb{B}) e^{-\beta_1 t_{\text{máx}}} + r_2(\mathbb{B}) e^{\lambda_1 t_{\text{máx}}} + r_3(\mathbb{B}) e^{\lambda_2 t_{\text{máx}}}. \\ \frac{\partial G_3(\mathbb{B})}{\partial \beta_1} &= \frac{A_0}{\lambda_1 - \lambda_2} \left[ S(\mathbb{B}) + \beta_1 \frac{\partial S(\mathbb{B})}{\partial \beta_1} \right], \\ \frac{\partial S(\mathbb{B})}{\partial \beta_1} &= e^{-\beta_1 t_{\text{máx}}} \left[ \frac{\partial r_1(\mathbb{B})}{\partial \beta_1} + r_1(\mathbb{B}) \left( -t_{\text{máx}} - \beta_1 \frac{\partial G_2(\mathbb{B})}{\partial \beta_1} \right) \right] \\ &\quad + e^{\lambda_1 t_{\text{máx}}} \left[ \frac{\partial r_2(\mathbb{B})}{\partial \beta_1} + r_2(\mathbb{B}) \lambda_1 \frac{\partial G_2(\mathbb{B})}{\partial \beta_1} \right] \\ &\quad + e^{\lambda_2 t_{\text{máx}}} \left[ \frac{\partial r_3(\mathbb{B})}{\partial \beta_1} + r_3(\mathbb{B}) \lambda_2 \frac{\partial G_2(\mathbb{B})}{\partial \beta_1} \right]. \end{aligned}$$

$$\begin{aligned} \frac{\partial G_3(\mathbb{B})}{\partial \beta_j} &= \frac{A_0 \beta_1}{(\lambda_1 - \lambda_2)} \left[ -\frac{\frac{\partial \lambda_1}{\partial \beta_j} - \frac{\partial \lambda_2}{\partial \beta_j}}{\lambda_1 - \lambda_2} S(\mathbb{B}) + \frac{\partial S(\mathbb{B})}{\partial \beta_j} \right], \quad j \geq 2. \\ \frac{\partial S(\mathbb{B})}{\partial \beta_j} &= e^{-\beta_1 t_{\text{máx}}} \left[ \frac{\partial r_1(\mathbb{B})}{\partial \beta_j} - r_1(\mathbb{B}) \beta_1 \frac{\partial G_2(\mathbb{B})}{\partial \beta_j} \right] \\ &\quad + e^{\lambda_1 t_{\text{máx}}} \left[ \frac{\partial r_2(\mathbb{B})}{\partial \beta_j} + r_2(\mathbb{B}) \left( \lambda_1 \frac{\partial G_2(\mathbb{B})}{\partial \beta_j} + t_{\text{máx}} \frac{\partial \lambda_1}{\partial \beta_j} \right) \right] \\ &\quad + e^{\lambda_2 t_{\text{máx}}} \left[ \frac{\partial r_3(\mathbb{B})}{\partial \beta_j} + r_3(\mathbb{B}) \left( \lambda_2 \frac{\partial G_2(\mathbb{B})}{\partial \beta_j} + t_{\text{máx}} \frac{\partial \lambda_2}{\partial \beta_j} \right) \right] \quad j \geq 2. \end{aligned}$$

4.  $G_4(\mathbb{B}) = T_k$ ,  $0 < k < 1$  siendo  $T_k$  el primer tiempo donde la concentración en el compartimento  $C$  alcanza un  $100k\%$  de su concentración máxima. Es decir,  $T_k$  satisface:

$$\eta_C(T_k; \mathbb{B}) = k\eta_C(t_{\text{máx}}; \mathbb{B}),$$

equivalente a que:

$$r_1(\mathbb{B}) \left[ e^{-\beta_1 T_k} - k e^{-\beta_1 t_{\text{máx}}} \right] + r_2(\mathbb{B}) \left[ e^{\lambda_1 T_k} - k e^{\lambda_1 t_{\text{máx}}} \right] + r_3(\mathbb{B}) \left[ e^{\lambda_2 T_k} - k e^{\lambda_2 t_{\text{máx}}} \right] = 0. \quad (\text{B.3})$$

Al derivar implícitamente la ecuación B.3 con respecto a cada componente de  $\mathbb{B}$ , se obtienen las siguientes expresiones:

$$\begin{aligned} S_1(\mathbb{B}) &= -\beta_1 r_1(\mathbb{B}) e^{-\beta_1 T_k} + \lambda_1 r_2(\mathbb{B}) e^{\lambda_1 T_k} + \lambda_2 r_3(\mathbb{B}) e^{\lambda_2 T_k} \\ S_2(\mathbb{B}) &= e^{-\beta_1 T_k} \left( T_k r_1(\mathbb{B}) - \frac{\partial r_1(\mathbb{B})}{\partial \beta_1} \right) + k e^{-\beta_1 t_{\text{máx}}} \left( \frac{\partial r_1(\mathbb{B})}{\partial \beta_1} - \left( t_{\text{máx}} + \beta_1 \frac{\partial G_2(\mathbb{B})}{\partial \beta_1} \right) r_1(\mathbb{B}) \right) \\ &\quad + k e^{\lambda_1 t_{\text{máx}}} \left( \frac{\partial r_2(\mathbb{B})}{\partial \beta_1} + r_2(\mathbb{B}) \lambda_1 \frac{\partial G_2(\mathbb{B})}{\partial \beta_1} \right) + k e^{\lambda_2 t_{\text{máx}}} \left( \frac{\partial r_3(\mathbb{B})}{\partial \beta_1} + r_3(\mathbb{B}) \lambda_2 \frac{\partial G_2(\mathbb{B})}{\partial \beta_1} \right) \\ &\quad - e^{\lambda_1 T_k} \frac{\partial r_2(\mathbb{B})}{\partial \beta_1} - e^{\lambda_2 T_k} \frac{\partial r_3(\mathbb{B})}{\partial \beta_1} \end{aligned}$$

$$\begin{aligned}
 S_3(\mathbb{B}) &= -e^{-T_k \beta_1} \frac{\partial r_1(\mathbb{B})}{\partial \beta_j} + k e^{-\beta_1 t_{\max}} \left( \frac{\partial r_1(\mathbb{B})}{\partial \beta_j} - \beta_1 r_1(\mathbb{B}) \frac{\partial G_2(\mathbb{B})}{\partial \beta_j} \right) \\
 &\quad + k e^{\lambda_1 t_{\max}} \left[ \frac{\partial r_2(\mathbb{B})}{\partial \beta_j} + r_2(\mathbb{B}) \left( \lambda_1 \frac{\partial G_2(\mathbb{B})}{\partial \beta_j} + t_{\max} \frac{\partial \lambda_1}{\partial \beta_j} \right) \right] \\
 &\quad + k e^{\lambda_2 t_{\max}} \left[ \frac{\partial r_3(\mathbb{B})}{\partial \beta_j} + r_3(\mathbb{B}) \left( \lambda_2 \frac{\partial G_2(\mathbb{B})}{\partial \beta_j} + t_{\max} \frac{\partial \lambda_2}{\partial \beta_j} \right) \right] \\
 &\quad + e^{\lambda_1 T_k} \left( -\frac{\partial r_2(\mathbb{B})}{\partial \beta_j} - r_2(\mathbb{B}) T_k \frac{\partial \lambda_1}{\partial \beta_j} \right) + e^{\lambda_2 T_k} \left( -\frac{\partial r_3(\mathbb{B})}{\partial \beta_j} - r_3(\mathbb{B}) T_k \frac{\partial \lambda_2}{\partial \beta_j} \right), \quad j \geq 2.
 \end{aligned}$$

$$\frac{\partial G_4(\mathbb{B})}{\partial \beta_1} = \frac{\partial t_{\max}}{\partial \beta_1} = \frac{S_2(\mathbb{B})}{S_1(\mathbb{B})}$$

$$\frac{\partial G_4(\mathbb{B})}{\partial \beta_j} = \frac{\partial t_{\max}}{\partial \beta_j} = \frac{S_3(\mathbb{B})}{S_1(\mathbb{B})}, \quad j \geq 2.$$

# Bibliografía

- Atkinson, A. C. (2008), “DT-optimum designs for model discrimination and parameter estimation”, *Journal of Statistical Planning and Inference*, **138**, 56-64.
- 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. (1975a), “The designs of experiments for discriminating between two rival models”, *Biometrika*, **62**, 57-70.
- Atkinson, A. C.; Fedorov, V. V. (1975b), “Optimal design: Experiments for discriminating between several models”, *Biometrika*, **62**(2), 289-303.
- Berger, R. L.; Hsu, J. C. (1996) “Bioequivalence trials, intersection-union tests and equivalence confidence sets”, *Statistical Science*, **11**(4), 283–302.
- Biswas, A.; Chaudhuri, P. (2002), “An efficient design for model discrimination on parameter estimation in linear models”, *Biometrika*, **89**(3), 709-718.
- Borth, D. M. (1975), “A total entropy criterion for the dual problem of model discrimination and parameter estimation”, *Journal of the Royal Statistical Society. Series B (Methodological)*, **37** (1), 77–87.
- Box, G. E. P.; Lucas, H. L. (1959), “Design of experiments in nonlinear situations”, *Biometrika*, 77-90.
- Brown, L. D.; Olkin I.; Sacks, J.; Wynn, H. P.(Eds) (1985), *Jack Karl Kiefer Collected Papers III, Design of Experiments*, Springer Verlag, New York.
- Candas, B., Lalonde, J. Normand, M. (1988), “A model of the distribution and metabolism or corticotropin-releasing factor”, *American Physiological Society*, 104-112.

- Chaloner, K.; Verdine, I. (1995), “Bayesian experimental design: A review”, *Statistical Science*, **10**(3), 273-304.
- Chernoff, H. (1953), “Locally optimal designs for estimating parameters”, *The Annals of Mathematical Statistics*, **24**(24), 586–602.
- Clark, B. y Smith, D. A. (1989), *An Introduction to Pharmacokinetics*, Oxford.
- Clyde, M. ; Chaloner, K. (1996), “The equivalence of constrained and weighted designs in multiple objective design problems”, *Journal of the American Statistical Association*, **91**(435), 1236–1244.
- Cochran, W. G. (1973), “Experiments for nonlinear functions”, *Journal of the American Statistical Association*, **68**(344), 771–781.
- Davis, J. L.; Papich, M. G.; Morton, A. J.; Gayle, J. Blikslager, A. T.; Campbell, N. B. (2007), “Pharmacokinetics of etodolac in the horse following oral and intravenous administration”, *J. vet. Pharmacol. Therap*, **30**, 43–48.
- Dette, H., Haines, L. M.; Imhof, L. A. (2007), “Maximin and Bayesian optimal designs for regression models”, *Statistica Sinica*, **17**, 463–480.
- Dette, H.; Melas, V. B.; Pepelyshev, A. (2004a), “Optimal designs for estimating individual coefficients in polynomial regression -a functional approach”. *Journal of Statistical Planning and Inference*. **118**, 201-219.
- Dette, H.; Melas, V. B.; Pepelyshev, A. (2004b), “Optimal designs for a class of nonlinear regression models”, *The Annals of Statistics*, **32**(5), 2142-2167.
- Dette, 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.
- Dette, H.; Sahm, M. (1998), “Minimax optimal designs in nonlinear regression models”, *Statistica Sinica*, **8**, 1249–1264.
- Diamon, T.; Goto, M. (2007), “Curvature-adjusted optimal design of sampling times for the inference of pharmacokinetic compartment models”, *Statistics in Medicine*, **25**, 2799–2812.
- Fedorov, V. V.; Hackl, P. (1997), *Model-Oriented Design of Experiments*, Lecture Notes in Statistics, New York.
- Fedorov, V. V.; Khabarov, V. (1986), “Duality of optimal designs for model discrimination and parameter estimation”, *Biometrika*, **73** (1), 183–190.
- Felsenstein, K. (1992), “Optimal Bayesian design for discrimination among rival models”, *Comput. Stat. Data An.* , **14** (4), 427–436.



- Ford, I.; Torsney, 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** (2), 569-583.
- Glen, I. (2005), “Pharmacokinetic analysis”, *Anaesthesia and Intensive Care Medicine*, **6**(8), 280–282.
- Gueorguieva, I.; Ogungbenrob, K.; Grahamc, G.; Glatta, S.; Aaronso, L. (2007), “A program for individual and population optimal design for univariate and multivariate response pharmacokinetic pharmacodynamic models”, *Computer Methods and Programs in Biomedicine*, **86**, 51–61.
- Gueorguieva, I.; Aaronso, L.; Ogungbenrob, K.; Jorga, K. M.; Rodgers, T.; Rouland, M. (2006), “Optimal design for multivariate response pharmacokinetic pharmacodynamic models”, *Journal of Pharmacokinetics and Pharmacodynamics*, **33** (2), 97–124.
- Hoel, P. G.; Levine, A. (1964), “Optimal spacing and weighting in polynomial prediction”, *The Annals of Mathematical Statistics*, **35**(4), 1553–1560.
- Huang, Y. C.; Wong, W. K. (1998), “Sequential construction of multiple-objective optimal designs”, *Biometrics*, **54**(4), 1388–1397.
- Jennrich, R. L.; Bright, P. B. (1976), “Fitting systems of linear differential equations using computer generated exact derivatives (with discussion)”, *Technometrics*, **18**(4), 385–399.
- Karlin, S.; Studden, W. J. (1966), “Optimal experimental designs”, *The Annals of Mathematical Statistics*, **37**(4), 783-815.
- Kiefer, J. (1959), “Optimum experimental designs (with discussion)”, *Journal Royal Statistical Society, B*, **21**, 272-319.
- Kiefer, J.; Wolfowitz, J. (1959), “Optimum designs in regression problems”, *The Annals of Mathematical Statistics*, **30**(2), 271-294.
- Kiefer, J.; Wolfowitz, J. (1960), “The equivalence of two extremum problems”, *Canadian Journal of Mathematics*, **12**, 363-366.
- Kiefer, J. (1974), “General equivalence theory for optimum designs (Approximate theory)”, *The Annals of Statistics*, **2** (5), 849–879.
- López-Fidalgo, J.; Rodríguez-Díaz, J. M., Sánchez, G. y Santos-Martín, M. T. (2005), “Optimal designs for compartmental models with correlated observations”, *Journal of Applied Statistics*, **32**(10), 1075–1088.
- López-Fidalgo, J.; Tommasi, C.; Trandafir, P. C. (2007), “An optimal experimental design criterion for discriminating between non-normal models”, *Journal of The Royal Statistical Society Series B.*, **69**(2), 231–242.

- López-Fidalgo, J. y Wong, W. K. (2002), “Design issues for the Michaelis-Mentel model”, *Journal Theoretical of Biology*, **215**(1), 1–11.
- López, R. V.; Ramos, Q. R. (2007a), “Una introducción a los diseños óptimos”, *Revista Colombiana de Estadística*, 1, 37–51.
- López, R. V.; Ramos, Q. R. (2007b), “Diseños experimentales óptimos en modelos de compartimientos”, *Memorias del XXI Foro Nacional de Estadística*, Instituto Nacional de Estadística, Geografía e Informática, 149–153.
- López, R. V.; Ramos, Q. R. (2007c) “Bayesian L-optimal and  $D_A$ -optimal Designs for a Compartmental Model”, *Poster Session in: mODa 8*, Almagro, España.
- López, R. V.; Ramos, Q. R. (2008) “Optimum Designs for the Dual-Problem Discrimination and Nonlinear Functions Estimation for a Compartmental Model”, *Poster session in: 1st Canada-Mexico statistics meeting*, Guanajuato, México.
- Metzler, C. M. (1971), “Usefulness of the two-compartmental open model in pharmacokinetics”, *Journal of the American Statistical Association*, **66** (333), 49-53.
- Müller, W. G. ; Ponce de Leon, A. C. M. (1996), “Discrimination between two binary data models: sequentially designed experiments”, *Journal of Statistical Computation and Simulation*, **55**, 87–100.
- O’Brien, T. E. (1992), “A note on quadratic designs for nonlinear regression models”, *Biometrika*, **79**(4), 847–849.
- O’Brien, T. E. (2005), “Designing for parameter subsets in Gaussian nonlinear regression models”, *Journal of Data Science*, **3**, 179–197.
- O’Brien, T. E.; Rawlings, J. O (1996), “A nonsequential design procedure for parameter estimation and model discrimination in nonlinear regression models”, *Journal of Statistical Planning and Inference*, **55**, 77–93.
- Pepelyshev, A. (2007), “Optimal designs for the exponential model with correlated observations”, *En: mODa 8-Adeances in Model-Oriented Design and Analysis*, Physica-Verlang, Heilderberg, New York, 165–172.
- Ponce de Leon, A. C.; Atkinson, A. C. (1991), “Optimum experimental design for discriminating between two Rival models in the presence of prior information”, *Biometrika*, **78**(3), 601–608.
- Ponce de Leon, A. C.; Atkinson, A. C. (1991), “The design of experiments to discriminate between two rival generalized linear models”, *En: Advances in GLIM and Statistical Modelling, vol. 78 of Lecture Notes in Statistics*, L. Fahrmeir, B. Francis, R. Gilchrist, and G. Tutz, eds., *Biometrika*, 159–164, Springer-Verlag, New York.
- Pukelsheim, F. (1993), *Optimal Design of Experiments*, John Wiley and 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 A language and environment for statistical computing (2005), *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL, <http://www.R-project.org>.
- Rajaram, L.; Tsokos, C. P. (2006), “Statistical modelling of a pharmacokinetic system”, *Stochastic Analysis and Applications*, **24**: 1061-1081.
- Seber, G. A. F.; Wild, C. J. (1989), *Nonlinear Regression*, John Wiley & Sons, New York.
- Silvey, S. D.; Titterington, D. M. (1974), “A langragian approach to optimal design”, *Biometrika*, **61**(2), 299–302.
- 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.
- Spezzaferri, F. (1988), “Nonsequential designs for model discrimination and parameter estimation”, En *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds. ), 777–783. Oxford Univ. Press.
- Staub, J. F.; Foos, E.; Courtin B.; Jochemsen, R.; Perault-Staub, A. M. (2003), “A non-linear compartmental model of Sr metabolism. II. Its physiological relevance for Ca metabolism”, *Am. J. Physiol. Regulatory Integrative Comp Physiol.*, **284**(3): R835-R852.
- Stehlík, M.; Rodríguez-Díaz, J. M.; Müller, W. G.; López-Fidalgo, J. (2008), “Optimal allocation of bioassays in the case of parametrized covariance functions: an application in lung’s retention of radioactive particles”. *TEST*, **17**, 56–68.
- Stigler, S. M. (1971), “Optimal experimental design for polynomial regression”, *Journal of the American Statistical Association*, **66** (334), 311-318.
- Stroud, J. R.; Muller, P.; Rosner, G. L. (2001), “Optimal sampling times in population pharmacokinetic Studies”, *Applied Statistics*. **50** (3) 345-359.
- Ucinski, D.; Bogacka, B. (2005), “T-optimum designs for discrimination between two multiresponse dynamic models”, *Journal Royal Statistical Society, B*, **67**, Part 1, 3–18.
- Wald, A. (1943), “On the efficient design of statistical investigations”, *The Annals of Mathematical Statistics*, **14**, 134-140.
- Waterhouse, T. H.; Redmann, S.; Duffull, S. B.; Eccleston, J. A (2005), “Optimal design for model discrimination and parameter estimation for Itraconazole population pharmacokinetics in cystic fibrosis patients”, *Journal of Pharmacokinetics and Pharmacodynamics*, **32**(3-4), 521–545.

- Waterhouse, T. H.; Woods, D. C.; Eccleston, J. A; Lewis, S. M. (2008), “Design selection criteria for discrimination/estimation for nested models and a binomial response”, *Journal of Statistical Planning and Inference*, **138**, 132–144.
- Whittle, P. (1973), “Some general points in the theory of optimal experimental design”, *Journal Royal Statistical Society. Series B (Methodological)*, **35**(1), 123–130.
- Zen, M. M.; Tsai, M. H. (2002), “Some criterion-robust optimal designs for the dual problem of model discrimination and parameter estimation”, *Sankhya, Series B*, **64**(Pt 3), 322–338.

# Anexos

En esta última parte del trabajo se anexan cuatro contribuciones que se presentaron en diferentes eventos. Los dos primeros anexos fueron documentos editados de dos posters presentados en el congreso Moda 8, *Model Oriented Data Analysis* en junio de 2007; Almagro, España y en el evento: *First Canada and Mexico Statistics Meeting* en febrero de 2008, CIMAT, Guanajuato, Gto. México. En el primer trabajo se estudiaron diseños  $\mathbb{L}$ - y  $\mathbb{D}_A$ -óptimos para un primer modelo de compartimientos que inicialmente fue considerado. En el segundo trabajo se reportaron algunas de las aportaciones presentadas en el capítulo 4. Los últimos dos anexos son publicaciones hechas en la Revista Colombiana de Estadística, 2007, sobre una introducción a los diseños óptimos; y en las Memorias del XXI Foro Nacional de Estadística, 2007, donde se hace una exposición somera de los diseños experimentales óptimos en un modelo de compartimientos.

**Poster session in: Moda 8**  
**Bayesian  $L$ -optimal and  $D_A$ -optimal Designs for a**  
**Compartmental Model.**

---

## 1 Introduction

We are interested in a four compartment model. The compartmental model is usually used in pharmacokinetics. Pharmacokinetics is dedicated to the study of the time course of substances and their relationship with an organism. A pharmacokinetic model is used to describe the concentration of such substances into the organism over time. In this area, practitioners are usually interested in several pharmacokinetic parameters, namely the total clearance ( $CL_t$ ), the volume of distribution ( $V_d$ ) and the area under the curve ( $auc$ ), among others, which are related through nonlinear functions. For instance, the total clearance is related to the dose ( $Do$ ) and the area under the concentration curve through the function:  $Cl_t = \frac{Do}{auc}$ , see Clark and Smith (1989) and ?. Pharmacokinetic data is gathered for each subject over time, so the first issue is to define the optimal sampling times in order to estimate several characteristics in the compartmental model of interest, see Section 2. In this work, we find optimal experimental designs for a nonlinear model arising in the last compartment,  $C$ , (for similar works see Atkinson et al. (1993), Allen (1983), Stroud et al. (2001) and Waterhouse (2005)). The optimization is performed with respect to various criteria which depend on the Fisher information matrix. In Section 3 we will define the Bayesian  $L$ -optimality and Bayesian  $D_A$ -optimal criteria (generalized determinant criterion) in order to obtain optimal sampling times for estimating several nonlinear functions of the parameters (rates of transfer) as the total area under the concentration curve, the maximum concentration and the time to maximum concentration (Atkinson and Donev (1992)). Several prior distributions have been used for both criteria and we have obtained all optimal designs by using the algorithms proposed in Fedorov and Hackl (1997), see Section 4. All efficiencies are provided and compared.

## 2 The Compartmental Model

We are interested in the following four compartment model, with reversible rates in the last two compartments: where  $\theta_i, i = 1, \dots, 5$  are called rates of transfer.

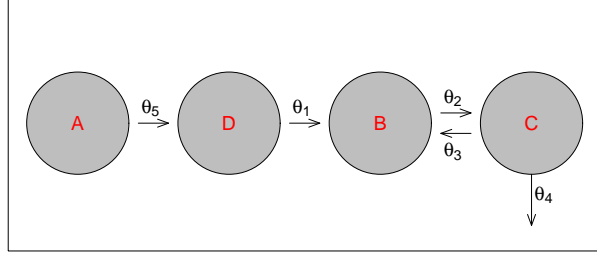


Figure 1: Four Compartment Model

This model has associated with the following linear system of ordinary differential equations:  $(\Theta^T = (\theta_1, \dots, \theta_5))$

$$\begin{aligned} \frac{d\eta_A(\Theta, t)}{dt} &= -\theta_5\eta_A(\Theta, t), \\ \frac{d\eta_D(\Theta, t)}{dt} &= \theta_5\eta_A(\Theta, t) - \theta_1\eta_D(\Theta, t), \\ \frac{d\eta_B(\Theta, t)}{dt} &= \theta_1\eta_D(\Theta, t) - \theta_2\eta_B(\Theta, t) + \theta_3\eta_C(\Theta, t), \\ \frac{d\eta_C(\Theta, t)}{dt} &= \theta_2\eta_B(\Theta, t) - (\theta_3 + \theta_4)\eta_C(\Theta, t), \end{aligned}$$

where  $\eta_{(*)}$  denotes the concentration in the compartment  $(*)$ . By solving the last system, we obtain the solution for the concentration of compartment  $C$ :

$$\eta_C(\Theta, t) = g_1(\Theta) \left[ g_2(\Theta)e^{-\theta_1 t} + g_3(\Theta)e^{-\theta_5 t} + g_4(\Theta)e^{k^{(1)} t} + g_5(\Theta)e^{k^{(2)} t} \right] \quad (1)$$

where:

$$\begin{aligned} g_1(\Theta) &= \frac{\theta_1 C_0 \theta_5 (k^{(2)} + \theta_2)(k^{(1)} + \theta_2)}{\theta_3 (\theta_5 - \theta_1)(k^{(1)} - k^{(2)})}, & g_2(\Theta) &= \frac{k^{(2)} - k^{(1)}}{(k^{(1)} + \theta_1)(k^{(2)} + \theta_1)}, \\ g_3(\Theta) &= \frac{k^{(1)} - k^{(2)}}{(k^{(1)} + \theta_5)(k^{(2)} + \theta_5)}, & g_4(\Theta) &= \frac{\theta_1 - \theta_5}{(k^{(1)} + \theta_1)(k^{(1)} + \theta_5)}, \\ g_5(\Theta) &= \frac{\theta_5 - \theta_1}{(k^{(2)} + \theta_1)(k^{(2)} + \theta_5)}, \end{aligned}$$

and:

$$k^{(1)} k^{(2)} = \theta_2 \theta_4, \quad k^{(1)} + k^{(2)} = \theta_2 + \theta_3 + \theta_4, \quad C_0 = \eta_A(\Theta, 0).$$

We consider the following characteristics of interest:

- Area under the curve:  $H_1(\Theta) = \int_0^\infty \eta_C(\Theta, t) dt$ .

- Time to maximum concentration:  $H_2(\Theta) = t_{\max} = \arg \max_t \eta_C(\Theta, t)$ .
- Maximum concentration:  $H_3(\Theta) = \eta_C(\Theta, t_{\max})$ .
- Total rates:  $H_4(\Theta) = \sum_{i=1}^5 \theta_i$ .
- Difference between reversible rates:  $H_5(\Theta) = \theta_2 - \theta_3$ .

Let  $\xi = \begin{pmatrix} t_0 & t_1 & \dots & t_k \\ w_0 & w_1 & \dots & w_k \end{pmatrix}$  be an approximate design where  $t_i \in [0, \infty)$ ,  $w_i > 0$  and  $\sum_i w_i = 1$ .

The purpose of this work is to obtain designs,  $\xi$ , that estimate all previous characteristics in an optimal way. In what follows,  $K(\Theta)$  denotes the matrix where the column  $j$  is the gradient function of  $H_j(\Theta)$ ,  $f(t, \Theta)$  is the gradient function associated with  $\eta_C(\Theta, t)$ , and  $M(\xi, \Theta)$  is the Fisher information matrix defined by  $M(\xi, \Theta) = \int f(t, \Theta) f^T(t, \Theta) d\xi(t)$ . We emphasize, in this case, that the optimal designs depend on a local value of  $\Theta$ . We are interested in finding optimal designs when some prior distribution for  $\Theta$  is known.

### 3 Bayesian $L$ -optimality and $D_A$ -optimality Criteria

We consider the following definitions:

1. The design  $\xi^*$  is a Bayesian  $L$ -optimal design with respect to the prior distribution  $\pi$ , abbreviated by  $L_\pi$ -optimal, if:

$$\begin{aligned} \xi^* &= \arg \min_{\xi} \mathbb{E}_{\pi} [\text{Tr} (K^T(\Theta) M^{-1}(\Theta, \xi) K(\Theta))] \\ &= \arg \min_{\xi} \int \text{Tr} (K^T(\Theta) M^{-1}(\Theta, \xi) K(\Theta)) \pi(\Theta) d\Theta \end{aligned}$$

It is known by applying the equivalence theorem, Fedorov and Hackl (1997), that:  $\xi^*$  is a  $L_\pi$ -optimal design if and only if:

$$\forall t \in [0, \infty) \quad \mathbb{E}_{\pi} [f^T(t) M^{-1}(\xi^*) K K^T M^{-1}(\xi^*) f(t) - \text{Tr} (K^T M^{-1}(\xi^*) K)] \leq 0.$$

2. The design  $\xi^*$  is a Bayesian  $D_A$ -optimal design with respect to a prior distribution  $\pi$ , abbreviated by  $D_{A\pi}$ -optimal if:

$$\begin{aligned} \xi^* &= \arg \min_{\xi} \mathbb{E}_{\pi} [-\log |A(\Theta) M^{-1}(\xi, \Theta) A^T(\Theta)|] \\ &= \arg \min_{\xi} \int -\log |A(\Theta) M^{-1}(\xi, \Theta) A^T(\Theta)| \pi(\Theta) d\Theta \end{aligned}$$

Again, from the equivalence theorem:  $\xi^*$  is a  $D_{A\pi}$ -optimal design if and only if:

$$\forall t \in [0, \infty) \quad \mathbb{E}_{\pi} \left[ f^T(t) M^{-1}(\xi^*) A^T (A M^{-1}(\xi^*) A^T)^{-1} A M^{-1}(\xi^*) f(t) \right] - k' \leq 0;$$

where  $k'$  is the number of rows of  $A$ . In our case,  $A = K^T$ .



In both cases  $f(t)$ ,  $M^-(\xi^*)$  and  $K$  are functions of  $\Theta$ . The expression on the left side is called the variance function.

## 4 Results

In order to analyze the effect of the prior distribution in the optimal design for both criteria, we consider three families of prior distributions. These families are a subset of all uniform discrete distributions defined in some appropriate subset of  $\mathbb{R}^5$ . They vary around some local value of  $\Theta$ ,  $\Theta_0$ .

1. Symmetric Prior Distribution, denoted by  $S(\delta)$ . The support of any member of this family contains  $\Theta_0$  and all combinations defined on:

$$\Omega_{I\delta} = \{\Theta \in \mathbb{R}^5 : \Theta \in (1 \pm \delta)\Theta_0\}$$

2. Left Asymmetric Prior Distribution, denoted by  $LA(\delta)$ . The support of any member of this family contains  $\Theta = (1 - \delta/2)\Theta_0$  and all combinations defined on:

$$\Omega_{II\delta} = \{\Theta_0\} \cup \{\Theta \in \mathbb{R}^5 : \Theta \in (1 - \delta)\Theta_0\}.$$

3. Right Asymmetric Prior Distribution, denoted by  $RA(\delta)$ . The support of this family contains  $\Theta = (1 + \delta/2)\Theta_0$  and all combinations defined on:

$$\Omega_{III\delta} = \{\Theta_0\} \cup \{\Theta \in \mathbb{R}^5 : \Theta \in (1 + \delta)\Theta_0\}.$$

In what follows, A prior distribution whose support is either  $\Omega_{I\delta}$  or  $\Omega_{II\delta}$  or  $\Omega_{III\delta}$  would be denoted by  $S\pi_\delta$ ,  $LA\pi_\delta$  or  $RA\pi_\delta$ , respectively.

We obtain optimal design numerically in the case when  $\Theta_0^T = (1.10 \ 2.82 \ 2.17 \ 0.54 \ 1.57)$ , and  $\delta \in \{0.10, 0.12, 0.14, 0.16\}$ . Also, we consider the prior distribution whose support is the set:  $\{\Theta_0\} \cup \bigcup_{\delta \in \{0.10, 0.12, 0.14, 0.16\}} \Omega_{I\delta}$ , abbreviated by  $\pi_{0.10} : \pi_{0.16}$ .

We use the algorithms proposed by Fedorov and Hackl (1997), with small changes in order to obtain the optimal designs. We check the optimality for each design by applying the equivalence theorem, see Figure 2. Figure 2 shows the variance functions associated with each optimal design for each prior distribution. Figure 3 shows the support points from optimal designs for all prior distributions.

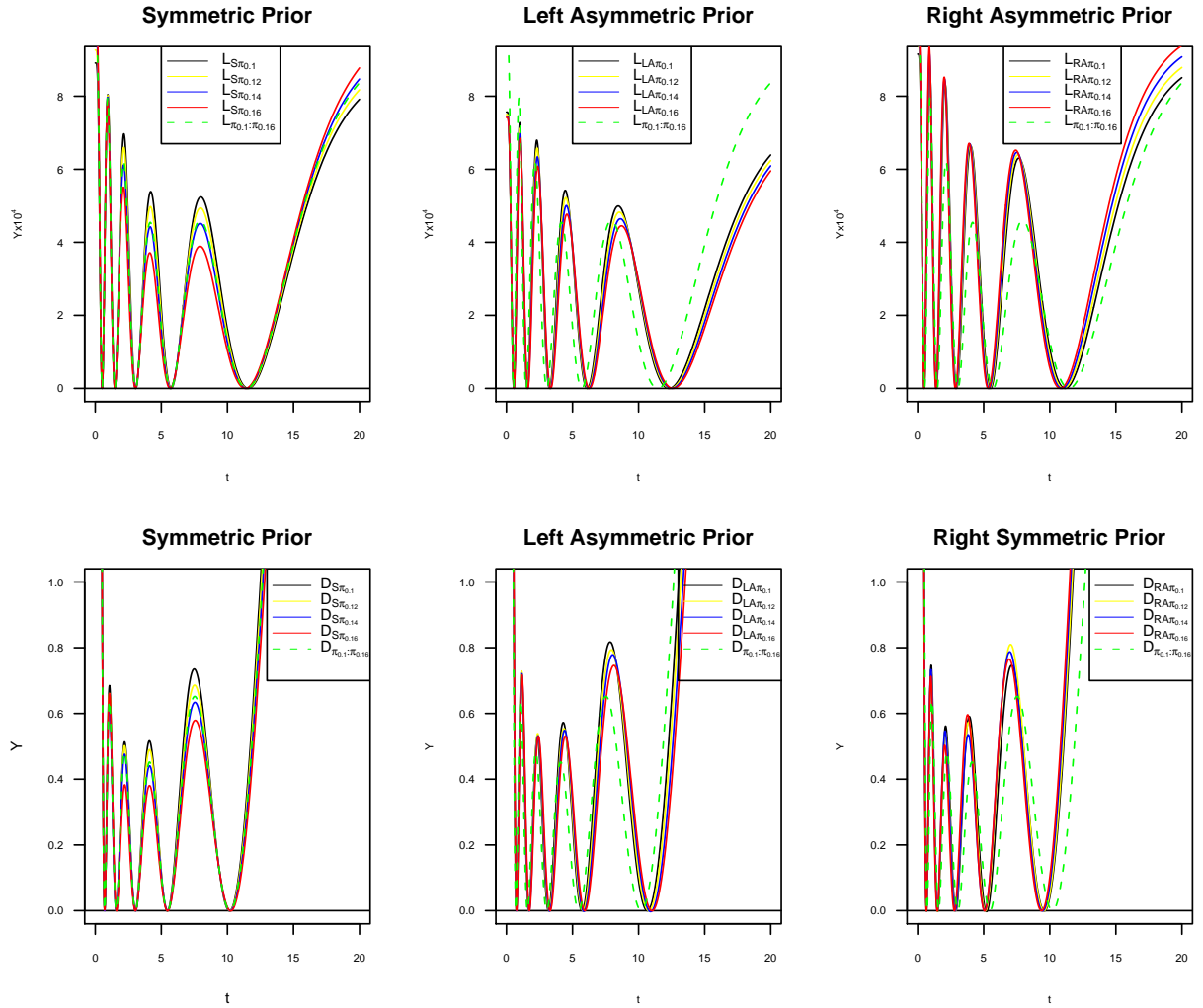


Figure 2: Variance functions for the optimal designs associated with Symmetric (S), Left Asymmetric (LA) and Right Asymmetric (RA) prior distributions by  $L_\pi$  and  $D_{A\pi}$ -optimality criteria.

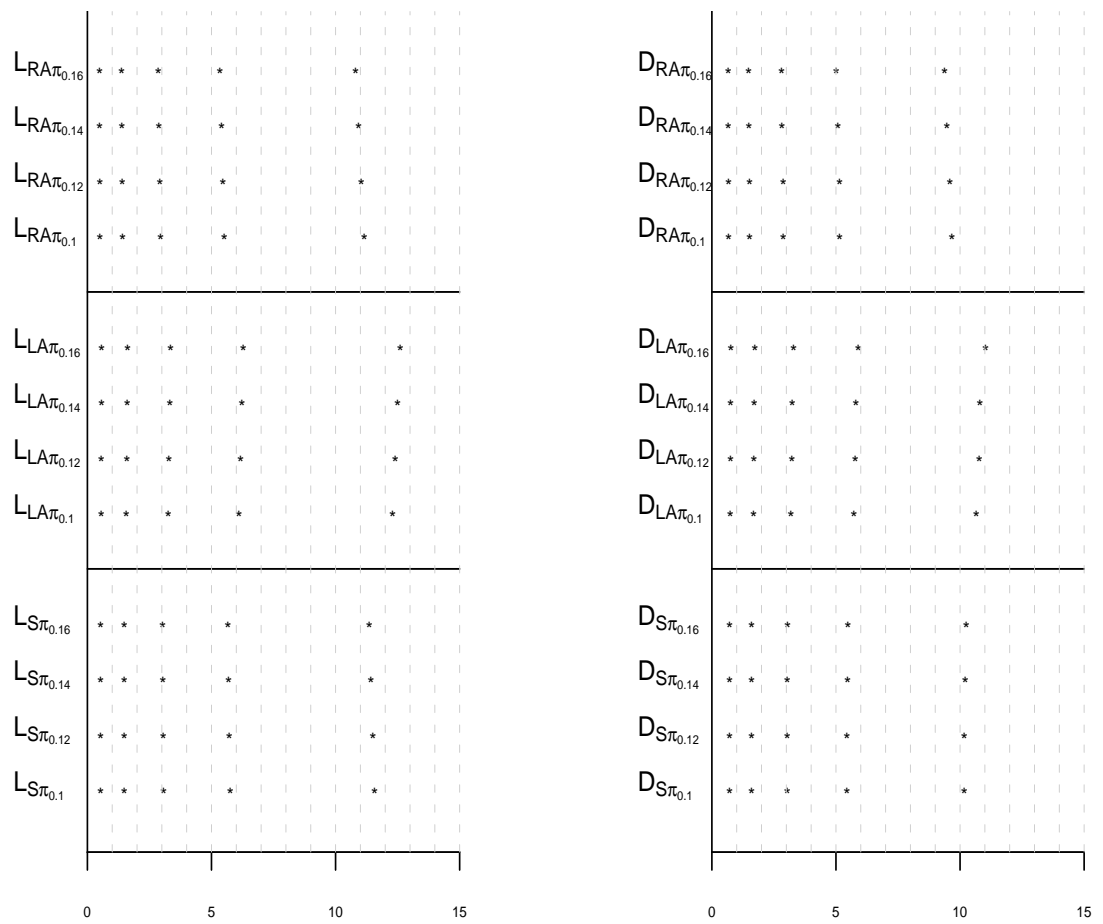


Figure 3: Support points associated to optimal designs for all prior distributions.

## 5 Efficiency

In order to compare both criteria we evaluate the efficiency of each criteria, defined by:

1. If  $\xi^*$  is  $L_\pi$ -optimal design, then the efficiency for  $L_\pi$ -optimality is:

$$e_{L_\pi}(\xi) = \frac{\mathbb{E}_\pi [\text{Tr} (K^T(\Theta)M^-(\xi^*, \Theta)K(\Theta))]}{\mathbb{E}_\pi [\text{Tr} (K^T(\Theta)M^-(\xi, \Theta)K(\Theta))]}.$$

2. If  $\xi^*$  is a  $D_{A\pi}$ -optimal design, then the efficiency for  $D_{A\pi}$ -optimality is:

$$e_{D_{A\pi}}(\xi) = \frac{\mathbb{E}_\pi [||K^T(\Theta)M^-(\xi^*, \Theta)K(\Theta)||]}{\mathbb{E}_\pi [||K^T M^-(\xi, \Theta)K||]}.$$

Table 2 shows the efficiencies for  $L_\pi$ -optimality and  $D_{A\pi}$ -optimality in the case when the support of  $\pi$  is defined in  $\Omega_{I\delta}$ ,  $\delta \in \{0.10, 0.12, 0.14, 0.16\}$ .

We note that the efficiencies for  $L_\pi$ -optimality are varying around 0.95 while the efficiencies for  $D_{A\pi}$ -optimality are varying around 0.65. The efficiencies for the other prior distributions are similar.

## 6 Concluding Remarks

We have found all optimal designs for both criteria, in the case when we have several prior distributions for the parameter  $\Theta$ .

- The optimal weights for all  $L_\pi$  optimal designs are similar, near 0.473, 0.194, 0.120, 0.096 and 0.117 respectively, while the support points have some changes, see Table 1.
- In  $L_\pi$  optimal designs, the support points are varied according to the type of prior distribution used, for example, they are more dispersed in left asymmetric distributions, in contrast with right asymmetric distributions. Similar conclusions are obtained in  $D_{A\pi}$ -optimal designs.
- Efficiencies for both criteria are robust with respect to the prior distribution.

There is no clear advantage among the  $L_\pi$  and  $D_{A\pi}$  optimal designs. The  $L_\pi$  optimal designs have higher efficiencies but the  $D_{A\pi}$  optimal designs spread the sampling weights more evenly with the added gain of potentially better data for testing lack of fit. In conclusion, it is difficult to decide what kind of design to use, it depends on the specific problem.

Prior Distribution	Optimal Design					Criterion value
$L_{S\pi_{0.10}}$	0.53082	1.49323	3.07715	5.75373	11.57869	89090.4003
	0.47274	0.19403	0.12095	0.09622	0.11606	
$L_{S\pi_{0.16}}$	0.52961	1.48518	3.03785	5.66455	11.35624	102363.7000
	0.47267	0.19303	0.11995	0.09623	0.11812	
$L_{\pi_{0.10}:\pi_{0.16}}$	0.53004	1.48917	3.05500	5.70407	11.45228	95622.1763
	0.47283	0.19356	0.12047	0.09615	0.11698	
$L_{LA\pi_{0.10}}$	0.55906	1.57493	3.26268	6.11268	12.29546	75664.1299
	0.47223	0.19387	0.12098	0.09647	0.116446	
$L_{LA\pi_{0.16}}$	0.57746	1.62559	3.35416	6.28472	12.60824	74179.0452
	0.47091	0.19359	0.12089	0.09670	0.11791	
$L_{RA\pi_{0.10}}$	0.50625	1.42605	2.95049	5.52205	11.16529	91512.1946
	0.47320	0.19473	0.12158	0.09607	0.11442	
$L_{LA\pi_{0.16}}$	0.49179	1.38450	2.85950	5.34594	10.80522	99493.3058
	0.47377	0.19468	0.12142	0.09593	0.11420	
$D_{S\pi_{0.10}}$	0.71468	1.60696	3.04074	5.45354	10.17686	29.7338
	0.20000	0.20003	0.20002	0.19996	0.19999	
$D_{S\pi_{0.16}}$	0.71571	1.60316	3.05069	5.48925	10.24924	29.9543
	0.20003	0.20002	0.19998	0.19999	0.19998	
$D_{LA\pi_{0.10}}$	0.75088	1.68460	3.19464	5.72290	10.65551	29.6327
	0.20000	0.20000	0.20000	0.20000	0.20000	
$D_{\pi_{0.10}:\pi_{0.16}}$	0.71467	1.61067	3.04470	5.46797	10.21263	29.7340
	0.19999	0.20002	0.19999	0.20000	0.20000	
$D_{LA\pi_{0.16}}$	0.77760	1.74168	3.29416	5.90875	11.04067	29.6997
	0.20001	0.19999	0.20000	0.20001	0.19999	
$D_{RA\pi_{0.10}}$	0.68126	1.52229	2.87722	5.14845	9.68297	29.6263
	0.20000	0.19992	0.20002	0.20003	0.20003	
$D_{RA\pi_{0.16}}$	0.66086	1.48342	2.81585	5.02421	9.38277	29.6707
	0.19997	0.19997	0.20003	0.20003	0.20000	

Table 1: Optimal design for all prior distributions.

Criterion	Prior 0.10		Prior 0.12		Prior 0.14		Prior 0.16		Prior 0.10 : 0.16	
$L_{\pi}$ - Optimal	1.000	0.957	1.000	0.958	1.000	0.958	1.000	0.959	1.000	0.958
$D_{A\pi}$ - Optimal	0.649	1.000	0.653	1.000	0.658	1.000	0.662	1.000	0.657	1.000

Table 2: Efficiencies of  $L_{\pi}$ -optimal designs and  $D_{A\pi}$ -optimal designs under symmetric prior distributions.

## References

- D. M. Allen. Parameter estimation for nonlinear models with emphasis on compartmental models. *Biometrics*, 39:629–637, 1983.
- A. C. Atkinson, K. Chaloner, A. M. Herzberg, and J. Juritz. Optimum experimental designs for properties of a compartmental model. *Biometrics*, 49 (2):325–337, 1993.
- A. C. Atkinson and A. N. Donev. *Optimum Experimental Designs*. Oxford University Press, New York, 1992.
- B. Clark and D. A. Smith. *An Introduction to Pharmacokinetics*. Oxford, England, 1989.
- V. V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments*. Springer-Verlag, New York, 1997.
- J. R. Stroud, P. Müller, and G. L. Rosner. Optimal sampling times in population pharmacokinetic studies. *Appl. Statist.*, 50 Part 3:345–359, 2001.
- T. H. Waterhouse. *Optimal Experimental Design for Nonlinear and Generalised Linear Models*. PhD thesis, University of Queensland, Australia, 2005.

**Poster Session in: First Canada and Mexico  
Statistics Meeting.  
Optimum Designs for the Dual-Problem  
Discrimination and Nonlinear Functions Estimation  
for a Compartmental Model.**

---

## 1 Introduction

The compartmental model is usually used in pharmacokinetics. Pharmacokinetics is dedicated to the study of the time course of substances and their relationship with an organism. A pharmacokinetic model is used to describe the concentration of such substances into the organism over time. In this area, practitioners are usually interested in several pharmacokinetic parameters, namely the total clearance ( $CL_t$ ), the volume of distribution ( $V_d$ ) and the area under the curve ( $AUC$ ), among others, which are related through nonlinear functions. For instance, the total clearance is related to the dose ( $Do$ ) and the area under the concentration curve through the function:  $Cl_t = \frac{Do}{AUC}$ , see Clark and Smith (1989) and Eyzaguirre (2006). Pharmacokinetic data is gathered for each subject over time, so the first issue is to define the optimal sampling times in order to estimate several characteristics in the compartmental model of interest, see Section 2.

In this work, we are interested in two compartmental models, one with four components and reversible rates of transfer and the other with three components ( see Atkinson et al. (1993), Allen (1983), Stroud et al. (2001) and Waterhouse (2005) for similar works). The purpose is to find optimal experimental designs for a nonlinear model arising in a particular compartment,  $C$ , Figure 1. Optimum designs for a discrimination between models often have very poor properties for estimation of the parameters in the chosen model, Atkinson (2008). Here we use optimum design theory to provide designs of known properties with a specified balance between estimation and discrimination. The objective is obtain an optimal design in order to discriminate between both models and estimate in optimal way several nonlinear functions of the parameters (rates of transfer) as the total area under the concentration curve, the maximum concentration, the time to maximum concentration and the first time where the concentration attains the 50% of the maximum concentration (Atkinson and Donev (1992)). The optimization is performed with respect to various criteria which depend on the Fisher information matrix. Special

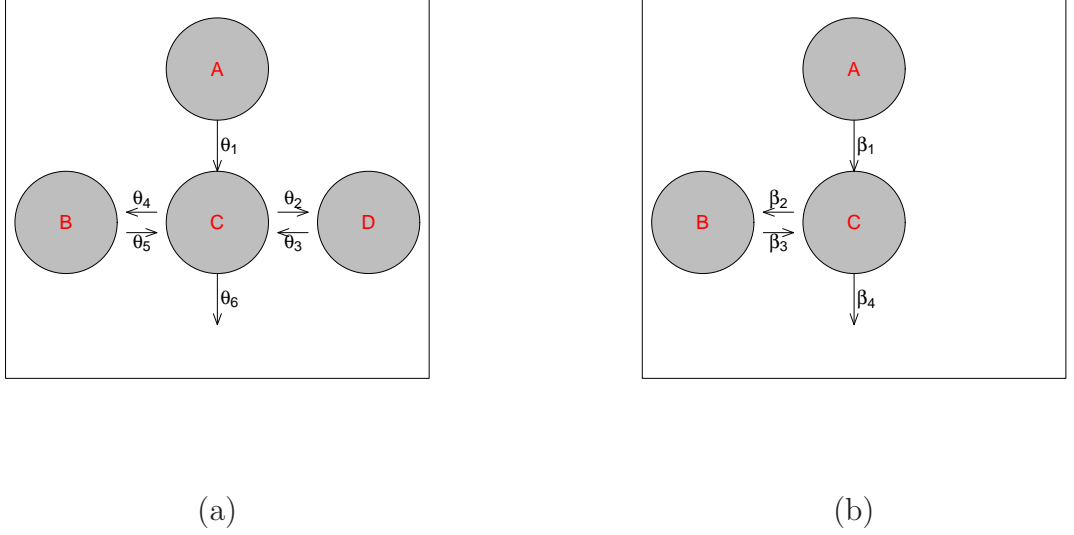


Figure 1: Two compartmental models, (a) Model *I*: Four-compartment model, with reversible rates in the last three compartments, with rates of transfer  $\theta_i, i = 1, \dots, 6$ , (b) Model *II*: Three-compartment model, with reversible rates in the last two compartments.

attention is given to Bayesian optimal design where several prior distributions are used and the optimal designs are obtained.

## 2 The Compartmental Models

We are interested in two compartmental models represented in Figure 1. The first model, Model *I*, has associated the following linear system of ordinary differential equations: ( $\Theta_1^T = (\theta_1, \dots, \theta_6)$ )

$$\begin{aligned}
 \frac{d\eta_A(t; \Theta_1)}{dt} &= -\theta_1 \eta_A(t; \Theta_1) \\
 \frac{d\eta_B(t; \Theta_1)}{dt} &= \theta_4 \eta_C(t; \Theta_1) - \theta_5 \eta_B(t; \Theta_1) \\
 \frac{d\eta_C(t; \Theta_1)}{dt} &= \theta_1 \eta_A(t; \Theta_1) - (\theta_2 + \theta_4 + \theta_6) \eta_C(t; \Theta_1) + \theta_5 \eta_B(t; \Theta_1) + \theta_3 \eta_D(t; \Theta_1) \\
 \frac{d\eta_D(t; \Theta_1)}{dt} &= \theta_2 \eta_C(t; \Theta_1) - \theta_3 \eta_D(t; \Theta_1),
 \end{aligned}$$

where  $\eta_{(*)}$  denotes the concentration in the compartment (\*). By solving the last system, we obtain the solution for the concentration of compartment *C*:

$$\eta_C^I(t; \Theta_1) = \theta_1 A_0 \left[ h_1(\Theta_1) e^{-\theta_1 t} + h_2(\Theta_1) e^{\lambda_1 t} + h_3(\Theta_1) e^{\lambda_2 t} + h_4(\Theta_1) e^{\lambda_3 t} \right], \quad (1)$$

where:

$$\begin{aligned}
 h_2(\Theta_1) &= \frac{(\lambda_1 + \theta_3)(\lambda_1 + \theta_5)}{(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)(\lambda_1 + \theta_1)}, & h_3(\Theta_1) &= -\frac{(\lambda_2 + \theta_3)(\lambda_2 + \theta_5)}{(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_2 + \theta_1)}, \\
 h_4(\Theta_1) &= \frac{(\lambda_3 + \theta_3)(\lambda_3 + \theta_5)}{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3)(\lambda_3 + \theta_1)}, & h_1(\Theta_1) &= -h_2(\Theta_1) - h_3(\Theta_1) - h_4(\Theta_1).
 \end{aligned}$$



and:

$\lambda_i$ ,  $i = 1, 2, 3$  are solutions of the cubic polynomial:

$$\lambda^3 + \lambda^2 (\theta_2 + \theta_3 + \theta_4 + \theta_5 + \theta_6) + \lambda ((\theta_2 + \theta_3 + \theta_6) \theta_5 + (\theta_4 + \theta_6) \theta_3) + \theta_3 \theta_5 \theta_6 = 0,$$

therefore the  $\lambda_i$ ,  $i = 1, 2, 3$ , satisfy:

1.  $-\sum_{i=1}^3 \lambda_i = \sum_{i=2}^6 \theta_i$ ,
2.  $\lambda_3(\lambda_1 + \lambda_2) + \lambda_1 \lambda_2 = (\theta_2 + \theta_3 + \theta_6) \theta_5 + \theta_3(\theta_4 + \theta_6)$ ,
3.  $-\lambda_1 \lambda_2 \lambda_3 = \theta_3 \theta_5 \theta_6$ .

The model *II* has associated the following linear system of ordinary differential equations:

$$\left( \Theta_2^T = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{pmatrix} \right)$$

$$\begin{aligned} \frac{d\eta_A(t; \Theta_2)}{dt} &= -\beta_1 \eta_A(t; \Theta_2) \\ \frac{d\eta_C(t; \Theta_2)}{dt} &= \beta_1 \eta_A(t; \Theta_2) + \beta_3 \eta_B(t; \Theta_2) - (\beta_4 + \beta_2) \eta_C(t; \Theta_2) \\ \frac{d\eta_B(t; \Theta_2)}{dt} &= \beta_2 \eta_C(t; \Theta_2) - \beta_3 \eta_B(t; \Theta_2). \end{aligned}$$

By solving the last system, we obtain the solution for the concentration of compartment *C*:

$$\eta_C^{II}(t; \Theta_2) = \frac{A_0 \beta_1}{\lambda_1 - \lambda_2} \left[ r_1(\Theta_2) e^{-\beta_1 t} + r_2(\Theta_2) e^{\lambda_1 t} + r_3(\Theta_2) e^{\lambda_2 t} \right] \quad (2)$$

where:

$$r_2(\Theta_2) = \frac{\lambda_1 + \beta_3}{\beta_1 + \lambda_1}, \quad r_3(\Theta_2) = -\frac{\lambda_2 + \beta_3}{\beta_1 + \lambda_2}, \quad r_1(\Theta_2) = -(r_2(\Theta_2) + r_3(\Theta_2)),$$

and,

$$\lambda_1 \lambda_2 = \beta_3 \beta_4, \quad \lambda_1 + \lambda_2 = -(\beta_2 + \beta_3 + \beta_4).$$

In both models, we consider the following characteristics of interest, which are nonlinear functions of  $\Theta$ :

- Area under the curve:  $H_1(\Theta_1^0) = \int_0^\infty \eta_C^I(t; \Theta_1^0) dt$  or  $G_1(\Theta_2^0) = \int_0^\infty \eta_C^{II}(t; \Theta_2^0) dt$ .
- Time to maximum concentration:  $H_2(\Theta_1^0) = t_{\max}^I = \arg \max_t \eta_C^I(t; \Theta_1^0)$  or  $G_2(\Theta_2^0) = t_{\max}^{II} = \arg \max_t \eta_C^{II}(t; \Theta_2^0)$ .
- Maximum concentration:  $H_3(\Theta_1^0) = \eta_C^I(t_{\max}^I; \Theta_1^0)$  or  $G_3(\Theta_2^0) = \eta_C^{II}(t_{\max}^{II}; \Theta_2^0)$ .
- First time to attains a 50% from maximum concentration:  $H_4(\Theta_1^0) = T_{0.5}^I$  or  $G_4(\Theta_2^0) = T_{0.5}^{II}$ ,  $T_{0.5}$  satisfies:  $\eta_C(T_{0.5}; \Theta) = 0.5 \eta_C(t_{\max}; \Theta)$ ,

where  $\Theta_1^0$  and  $\Theta_2^0$  are local values for  $\Theta_1$  and  $\Theta_2$  respectively.

Then, we consider the general nonlinear regression model:

$$Y(t) = \eta(t, \Theta) + \epsilon, \quad (3)$$

where the random variables  $\epsilon$  are independent and normally distributed with zero mean and constant variance  $\sigma^2$ . The function  $\eta(t, \Theta)$  may be one of the two partially known functions  $\eta_C^I(t; \Theta_1)$  and  $\eta_C^{II}(t; \Theta_2)$ , where  $\Theta_1 \in \Omega_1 \subseteq \mathbb{R}^6$ , and  $\Theta_2 \in \Omega_2 \subseteq \mathbb{R}^4$  are the unknown parameter vectors.

Let  $\xi = \begin{pmatrix} t_1 & t_2 & \dots & t_k \\ w_1 & w_2 & \dots & w_k \end{pmatrix}$  be an approximate design where  $t_i \in [0, \infty)$ ,  $w_i > 0$  and  $\sum_i w_i = 1$ .  $t_i$  represents the  $i$ -th sampling time and  $w_i$  the frequency of observations in this time.

The purpose of this work is to obtain optimal designs,  $\xi$ , that:

1. Discriminate between models  $I$  and  $II$  and
2. estimate the vector of parameters and all previous characteristics in an optimal way.

In what follows,  $K(\Theta)$  denotes the matrix where the column  $j$  is the gradient function of  $H_j(\Theta_1^0)$  or  $G_j(\Theta_2^0)$ ;  $f(t; \Theta)$  is the gradient function associated with  $\eta_C(t; \Theta)$ ; and the Fisher information matrix is defined by  $M(\xi; \Theta) = \int f(t; \Theta) f^T(t; \Theta) d\xi(t)$ . We emphasize, in this case, that the optimal designs depend on a local value of  $\Theta$ ,  $\Theta_1^0$  or  $\Theta_2^0$ . We are interested in finding local and Bayesian optimal designs when some prior distributions for  $\Theta$  are known. In the following we consider the most popular criteria,  $\mathbb{D}$ ,  $\mathbb{L}$ -optimality for estimating  $\Theta$  and the nonlinear functions, respectively; and to discriminate between both models. In the last case we use  $\mathbb{D}_s$  and  $\mathbb{T}$ -optimality. All criteria are concave real functions of the design  $\xi$ ,  $\phi(\xi)$ , called optimality criterion.

### 3 Optimum Designs for Discriminating and Estimating Nonlinear Functions Simultaneously (DENFS)

We define the following criterion as the result of considering the geometric mean of the efficiencies by the individuals criteria,  $\phi(\xi)/\max_\xi \phi(\xi)$  and then we take the logarithm of it. With  $0 \leq \alpha \leq 1$ , the criterion is:

$$\begin{aligned} \phi_\alpha(\xi; \Theta) = & \alpha \log \phi_1(\xi) + \frac{1-\alpha}{4} \left\{ \frac{1}{p_1} \log |M_1(\xi)| + \frac{1}{p_2} \log |M_2(\xi)| \right. \\ & \left. - \log [Tr (K_1^T M_1^{-1}(\xi; \Theta_1) K_1)] - \log [Tr (K_2^T M_2^{-1}(\xi; \Theta_2) K_2)] \right\}, \quad (4) \end{aligned}$$

where  $\Theta^T = (\Theta_1^T, \Theta_2^T)$  and  $\phi_1(\xi)$  denotes the discrimination  $\mathbb{D}_2$  or  $\mathbb{T}$ -optimality.

When  $\alpha = 1$ , we obtain the criterion for discriminating between models, and with  $\alpha = 0$  we obtain a criterion to estimate  $\Theta$  and nonlinear functions in both models simultaneously.

A design  $\xi$  is called local  $DENFS_\alpha$ -optimal if it maximizes 4. We denote this criterion by  $\phi_\alpha^T(\xi; \Theta)$  or  $\phi_\alpha^D(\xi; \Theta)$ , if  $\mathbb{T}$  or  $\mathbb{D}_2$ -optimality for discrimination, respectively. In  $\mathbb{T}$ -optimality case:

$\phi_T(\xi) = \min_{\Theta_2} \int \{\eta_1(t_i, \Theta_1^0) - \eta_2(t_i, \Theta_2)\}^2 d\xi$ , and  $\mathbb{D}_2$ -optimality:  $\phi_{D_2}(\xi) = \left\{ \frac{\det M_1(\xi)}{\det M_{11}(\xi)} \right\}^{1/2}$ .

The next theorem gives a necessary and sufficient condition for  $DENFS_\alpha$ -optimality.

**Theorem 3.1.** *A design  $\xi$  is a local  $DENFS_\alpha$ -optimal if and only if,  $\forall t \in [0, \infty)$ ,*

$$D(t; \xi, \alpha) = \alpha D_1(t; \xi) + \frac{1-\alpha}{4} \left\{ \frac{1}{p_1} f_1^T(t) M_1^{-1} f_1(t) + \frac{1}{p_2} f_2^T(t) M_2^{-1} f_2(t) + \frac{f_1^T(t) M_1^{-1} K_1 K_1^T M_1^{-1} f_1(t)}{\text{Tr}(K_1^T M_1^{-1} K_1)} + \frac{f_2^T(t) M_2^{-1} K_2 K_2^T M_2^{-1} f_2(t)}{\text{Tr}(K_2^T M_2^{-1} K_2)} \right\} - 1 \leq 0, \quad (5)$$

where  $D_1(t; \xi) = \frac{\Psi_1(t; \xi)}{\phi_T(\xi)}$  in  $\mathbb{T}$ -optimality case. And in  $\mathbb{D}_2$ -optimality case;  $D_1(t; \xi) = 0.5 * [f_1^T(t) M_1^{-1}(\xi) f_1(t) - f_{11}^T(t) M_{11}^{-1}(\xi) f_{11}(t)]$ .  $f_{11}^T$  has the components of the gradient except those associated with  $\theta_2$  and  $\theta_3$ ; and  $M_{11}(\xi)$  is the information matrix calculated with  $f_{11}$ .

The proof is an application of the known equivalence theorem in the theory of optimal designs. See Fedorov and Hackl (1997).

We consider local and Bayesian optimal designs, so that:

1. The design  $\xi^*$  is a local  $DENFS_\alpha$ -optimal design, if  $\xi^* = \arg \max_\xi \phi_\alpha(\xi; \Theta)$ .
2. The design  $\xi^*$  is a Bayesian  $DENFS_\alpha$ -optimal design with respect to a prior distribution  $\pi$  of  $\Theta$ , abbreviated by  $DENFS_{\alpha, \pi}$ -optimal, if:

$$\xi^* = \arg \max_\xi \mathbb{E}_\pi [\phi_\alpha(\xi; \Theta)] = \arg \max_\xi \int \phi_\alpha(\xi; \Theta) \pi(\Theta) d\Theta.$$

Again, from a similar equivalence theorem as 3.1:  $\xi^*$  is a  $DENFS_{\alpha, \pi}$ -optimal design if and only if:

$$\forall t \in [0, \infty), \quad \mathbb{E}_\pi \left[ \alpha D_1(t; \xi) + \frac{1-\alpha}{4} \left\{ \frac{1}{p_1} f_1^T(t) M_1^{-1} f_1(t) + \frac{1}{p_2} f_2^T(t) M_2^{-1} f_2(t) + \frac{f_1^T(t) M_1^{-1} K_1 K_1^T M_1^{-1} f_1(t)}{\text{Tr}(K_1^T M_1^{-1} K_1)} + \frac{f_2^T(t) M_2^{-1} K_2 K_2^T M_2^{-1} f_2(t)}{\text{Tr}(K_2^T M_2^{-1} K_2)} \right\} \right] - 1 \leq 0 \quad (6)$$

In both cases  $f_1(t)$ ,  $f_2(t)$ ,  $M_1^{-1}$ ,  $M_2^{-1}$  and  $K_1$ ,  $K_2$  are functions of  $\Theta$ . The expression on the left side is called the variance function.

## 4 Results

### 4.1 Local Designs

In this section we obtain locally  $DENFS(\alpha)$ -optimal designs in the cases where  $\mathbb{D}_2$  or  $\mathbb{T}$ -optimality criterion are used for discriminating between the models *I* and *II*. By illustration, we consider the local value of  $\Theta$ ,

$$\Theta_0 = [\Theta_1^0, \Theta_2^0] = [0.30 \ 0.20 \ 0.15 \ 0.05 \ 0.08 \ 0.25, \ 0.40 \ 0.28 \ 0.10 \ 0.30].$$

Case	$\alpha$	Optimal Design ( $\xi$ )						Optimal Value	
$C_1$	0.25	0.9322	3.7206	9.5987	19.5410	35.9549	67.5850	6.3384	
		0.1021	0.1123	0.1168	0.1511	0.2262	0.2914		
	0.50	0.8229	3.6806	9.6092	19.7227	36.3282	71.1595		6.7252
		0.1038	0.1081	0.1330	0.1548	0.2182	0.2821		
	0.75	0.7474	3.6810	9.6222	19.6769	36.3610	74.2231		7.0943
		0.1070	0.1041	0.1448	0.1593	0.2018	0.2830		
$C_2$	0.25	0.9783	3.7571	9.9886	21.8765	34.0914	62.6908	7.6925	
		0.0938	0.1045	0.1173	0.1994	0.0915	0.3935		
	0.50	0.8728	3.7070	10.0933	22.8589	35.8963	62.9366		9.4076
		0.0862	0.0911	0.1297	0.2262	0.0385	0.4283		
	0.75	0.7854	3.7187	10.1958	23.3111	37.7856	63.1089		11.0943
		0.0804	0.0779	0.1360	0.2345	0.0156	0.4556		

Table 1: Local  $DENFS(\alpha)$ -optimal designs, with  $\mathbb{D}_2$  optimality, ( $C_1$ ); and  $\mathbb{T}$  optimality, ( $C_2$ ) and different values of  $\alpha$ .

The optimal designs appear in the Table 1, with  $\alpha = 0.25, 0.50, 0.75$ . All designs satisfy the respectively equivalence theorem, Figure 2.

Also we analyze the  $\alpha$  selection in order to obtain a maximum product efficiency. We find the  $DENFS_\alpha$ -optimal design with  $\alpha = 0.01 : 0.01 : 0.99$  and we calculate the individuals efficiencies for all criteria considered, see Figure 3 (a), (b), and the product efficiencies associated with the  $DENFS_\alpha^D$ -optimality and  $DENFS_\alpha^T$ -optimality, Figure 3 (c). In this Figure, the best  $\alpha$  was 0.33, in both cases, and we concluding that  $DENFS_\alpha^D$ -optimality is better than  $DENFS_\alpha^T$  for all values of  $\alpha$ .

## 4.2 Bayesian Optimal Designs

In order to analyze the behavior of the prior distribution in the optimal design for both criteria, we consider three families of prior distributions. These families are a subset of all uniform discrete distributions defined in some appropriate subset of  $\mathbb{R}^p$ . They vary around some local value of  $\Theta$ ,  $\Theta_0$ .

1. Symmetric Prior Distribution, denoted by  $S_\delta$ . The support of any member of this family contains  $\Theta_0$  and all combinations defined on:

$$\Omega_{I\delta} = \{\Theta \in \mathbb{R}^p : \Theta \in (1 \pm \delta)\Theta_0\}$$

2. Left Asymmetric Prior Distribution, denoted by  $LA_\delta$ . The support of any member of this family contains  $\Theta = (1 - \delta/2)\Theta_0$  and all combinations defined on:

$$\Omega_{II\delta} = \{\Theta_0\} \cup \{\Theta \in \mathbb{R}^p : \Theta \in (1 - \delta)\Theta_0\}.$$

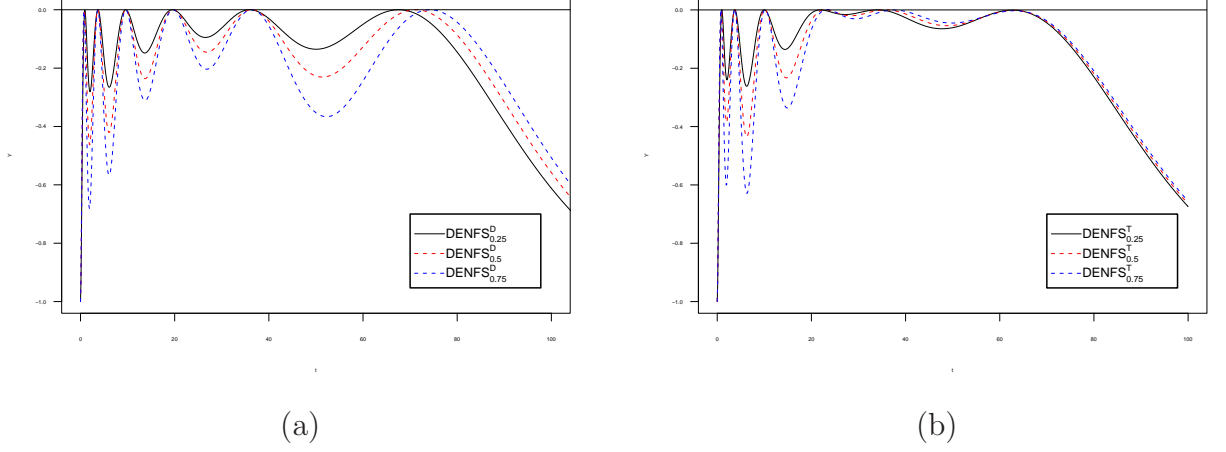


Figure 2: Variance functions for  $DENFS_\alpha$ -optimality,  $\alpha = 0.25, 0.50, 0.75$ , (a)  $\mathbb{D}_s$ -optimality, (b)  $\mathbb{T}$ -optimality.

3. Right Asymmetric Prior Distribution, denoted by  $RA_\delta$ . The support of this family contains  $\Theta = (1 + \delta/2)\Theta_0$  and all combinations defined on:

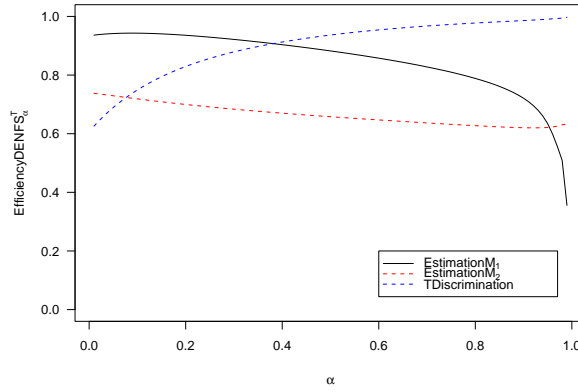
$$\Omega_{III\delta} = \{\Theta_0\} \cup \{\Theta \in \mathbb{R}^p : \Theta \in (1 + \delta)\Theta_0\}.$$

Two sets of prior distributions are considered:

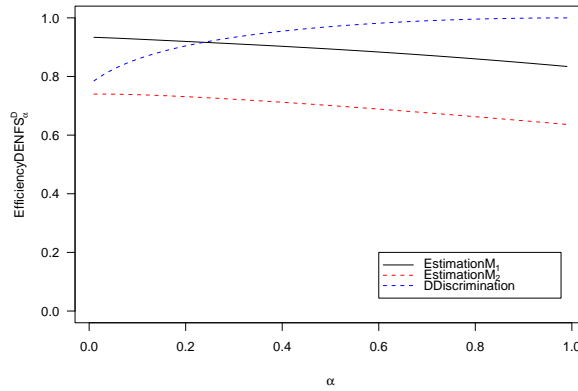
- $\pi_1$ : the support of this prior distribution is contained in  $\Omega_1 \times \{\Theta_2\}$ , where  $\Omega_1$  denotes the support associated with the prior distributions defined before in the case when  $\Theta = \Theta_1^0 \in \mathbb{R}^6$ . The criterion was denoted by  $DEFNS_{\alpha, S_\delta}^{*\pi}$  or  $DEFNS_{\alpha, LA_\delta}^{*,\pi}$  or  $DEFNS_{\alpha, RA_\delta}^{*\pi}$ , where  $*$  is  $\mathbb{D}$  or  $\mathbb{T}$ , if the criterion to discriminate is  $\mathbb{D}_2$  or  $\mathbb{T}$ -optimality, respectively.
- $\pi_2$ : the support of this prior is contained in  $\Omega$ ; the set of the support points associated with the prior distributions defined before where  $\Theta^{0T} = (\Theta_1^{0T}, \Theta_2^{0T}) \in \mathbb{R}^{10}$ .

We obtained the optimal designs numerically for  $\Theta_0^T$  and  $\delta \in \{0.05, 0.10, 0.15\}$ , and  $\alpha = 0.50$ . All designs were checked for optimality using the equivalence condition from equation 6, see Figure 4. Figure 4 shows the variance functions associated with each optimal design for each prior distribution.

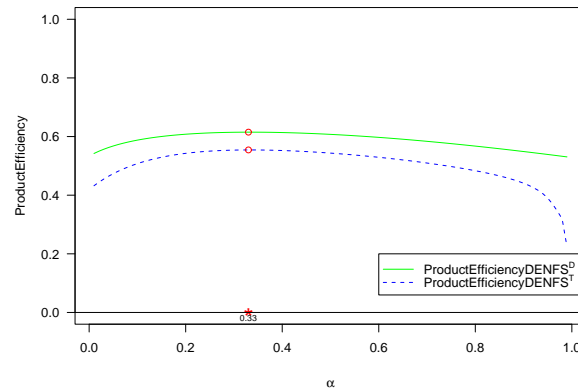
Figure 5 shows the support points from optimal designs for all prior distributions with both prior distributions,  $\pi_1$  and  $\pi_2$ . We observe that in both cases, the support points are very similar.



(a)



(b)



(c)

Figure 3: Efficiencies of designs as  $\alpha$  varies, (a)  $DENFS_{\alpha}^T$ , (b)  $DENFS_{\alpha}^D$ , (c) Product Efficiency.

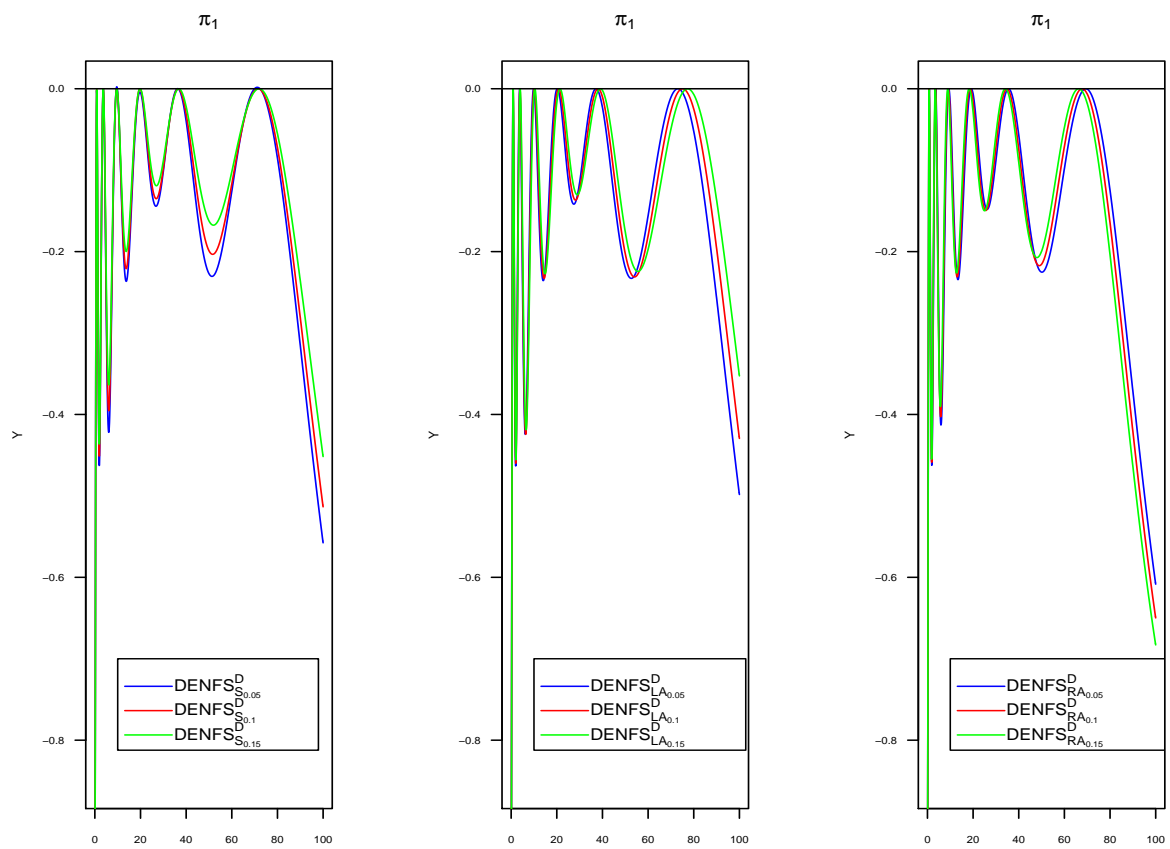


Figure 4: Joint plot of variance functions for bayesian *DENFS*-optimal designs, with prior  $\pi_1$  and  $\mathbb{D}_2$ -optimality for discriminating.

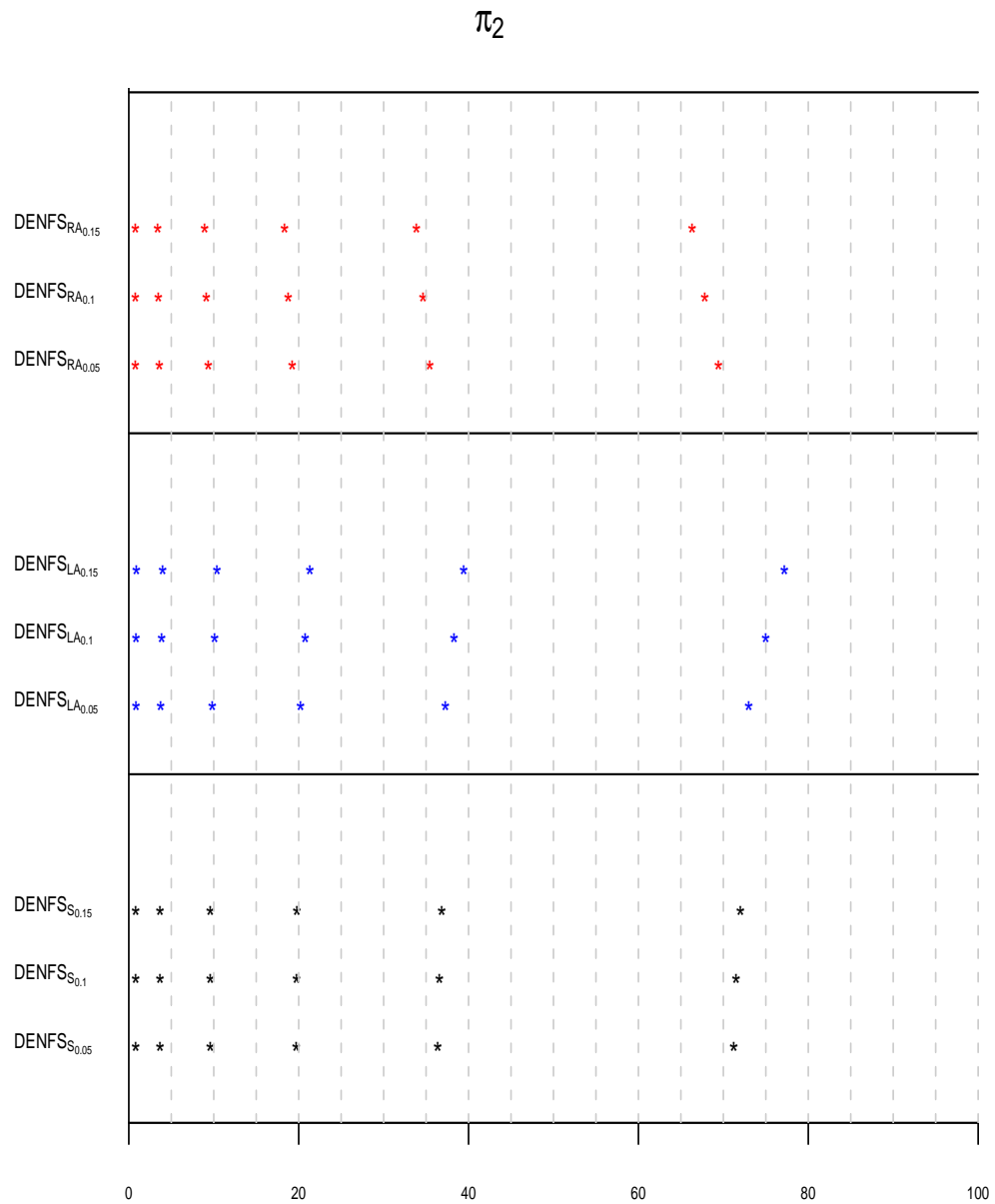


Figure 5: Joint plot of support points for the bayesian *DENFS* optimal design with  $\pi_2$ .



### 4.3 Efficiency

In order to compare both criteria, in the cases when we use  $\mathbb{T}$  and  $\mathbb{D}_2$ -optimality respectively, we evaluate the efficiency of each criterion, defined by:

1. Let  $\xi^*$  be a local  $DENFS_\alpha^D$ -optimal design, the efficiency for  $DENFS_\alpha^D$ -optimality is:

$$e_{DENFS_\alpha^D}(\xi) = \frac{\phi_\alpha^D(\xi)}{\phi_\alpha^D(\xi^*)}.$$

2. Let  $\xi^*$  be a local  $DENFS_\alpha^T$ -optimal design, the efficiency for  $DENFS_\alpha^T$ -optimality is:

$$e_{DENFS_\alpha^T}(\xi) = \frac{\phi_\alpha^T(\xi)}{\phi_\alpha^T(\xi^*)}.$$

3. Let  $\xi^*$  be a bayesian  $DENFS_\alpha^{\pi_1 D}$ -optimal design, the efficiency for  $DENFS_\alpha^{\pi_1 D}$ -optimality is:

$$e_{DENFS_\alpha^D}(\xi) = \frac{\mathbb{E}_{\pi_1} [\phi_\alpha^D(\xi; \Theta)]}{\mathbb{E}_{\pi_1} [\phi_\alpha^D(\xi^*, \Theta)]}.$$

In a similar way was defined in the case  $DENFS_\alpha^{\pi_1 T}$ .

Criterion	$\alpha = 0.25$		$\alpha = 0.50$		$\alpha = 0.75$	
local $DENFS_\alpha^D$ – optimal	1.0000	0.9839	1.0000	0.9470	1.0000	0.8844
local $DENFS_\alpha^T$ – optimal	0.9912	1.0000	0.9788	1.0000	0.9672	1.0000

Table 2: Efficiencies of local  $DENFS_\alpha^D$  – optimal designs and local  $DENFS_\alpha^T$  – optimal designs.

Table 2 shows the efficiencies for local  $DENFS_\alpha^D$ -optimal designs and local  $DENFS_\alpha^T$ -optimal, in the cases when  $\alpha = 0.25, 0.50, 0.75$ . For example, 0.9839 represents the efficiency from  $DENFS_{0.25}^T$ -optimal design with respect to  $DENFS_{0.25}^D$ -optimality criterion.

The efficiencies of  $DENFS_{0.5, \text{prior}}^D$  – optimal designs vary around 0.98 with respect to the criterion  $DENFS_{0.5, \text{prior}}^T$ , in all prior distribution considered, while the efficiencies of  $DENFS_{0.5, \text{prior}}^T$  – optimal designs vary between 0.94 and 0.96 with respect to the criterion  $DENFS_{0.5, \text{prior}}^D$ , see Table 3. The efficiencies are similar in all cases.

## 5 Concluding Remarks

- We define a criterion which is a balance between discrimination and nonlinear functions estimation.
- We use and compare two possible discrimination criteria. Also, in a local case, we analyze the behaviour of the efficiencies designs in function of  $\alpha$ .

Criterion	$S_{0.05}$	$S_{0.10}$	$S_{0.15}$
$DENFS_{0.5S_\delta}^D$ – optimal	1.0000 0.9481	1.0000 0.9558	1.0000 0.9640
$DENFS_{0.5S_\delta}^T$ – optimal	0.9797 1.0000	0.9821 1.0000	0.9854 1.0000
	$LA_{0.05}$	$LA_{0.10}$	$LA_{0.15}$
$DENFS_{0.5LA_\delta}^D$ – optimal	1.0000 0.9451	1.0000 0.9482	1.0000 0.9523
$DENFS_{0.5LA_\delta}^T$ – optimal	0.9790 1.0000	0.9798 1.0000	0.9811 1.0000
	$RA_{0.05}$	$RA_{0.10}$	$RA_{0.15}$
$DENFS_{0.5RA_\delta}^D$ – optimal	1.0000 0.9452	1.0000 0.9480	1.0000 0.9512
$DENFS_{0.5RA_\delta}^T$ – optimal	0.9790 1.0000	0.9796 1.0000	0.9805 1.0000

Table 3: Efficiencies of  $DENFS_{0.5,\text{prior}}^D$  – optimal designs and  $DENFS_{0.5,\text{prior}}^T$  – optimal designs, where prior is  $S_\delta$  (symmetric),  $LA_\delta$  (Left Asymmetric),  $RA_\delta$  (Right Asymmetric) with  $\delta = 0.05, 0.10, 0.15$ .

- In bayesian  $DENFS_{0.5,\text{prior}}^D$  optimal designs, the support points are varied according to the type of prior distribution used, for example, they are more dispersed in left asymmetric distributions, in contrast with right asymmetric distributions.
- Efficiencies for both criteria are robust with respect to the prior distribution and respect to the discrimination criteria

## References

- D. M. Allen. Parameter estimation for nonlinear models with emphasis on compartmental models. *Biometrics*, 39:629–637, 1983.
- A. C. Atkinson. DT-optimum designs for the Model Discrimination and Parameter Estimation. *Journal of Statistical Planning and Inference*, 138:56–64, 2008.
- A. C. Atkinson, K. Chaloner, A. M. Herzberg, and J. Juritz. Optimum experimental designs for properties of a compartmental model. *Biometrics*, 49 (2):325–337, 1993.
- A. C. Atkinson and A. N. Donev. *Optimum Experimental Designs*. Oxford University Press, New York, 1992.
- B. Clark and D. A. Smith. *An Introduction to Pharmacokinetics*. Oxford, England, 1989.
- R. H. Eyzaguirre. *Analysis and Design of Pharmacokinetic Models*. Master thesis, Technische Universiteit Eindhoven, Nethelards.
- V. V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments*. Springer-Verlag, New York, 1997.
- J. R. Stroud, P. Müller, and G. L. Rosner. Optimal sampling times in population pharmacokinetic studies. *Appl. Statist.*, 50 Part 3:345–359, 2001.
- T. H. Waterhouse. *Optimal Experimental Design for Nonlinear and Generalised Linear Models*. PhD thesis, University of Queensland, Australia, 2005.

# Una introducción a los diseños óptimos

## An Introduction to Optimal Designs

VÍCTOR IGNACIO LÓPEZ<sup>1,2,a</sup>, ROGELIO RAMOS<sup>2,b</sup>

<sup>1</sup>ESCUELA DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DE COLOMBIA, MEDELLÍN

<sup>2</sup>CENTRO 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:

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

---

<sup>a</sup>Profesor asistente. E-mail: vilopez@unalmed.edu.co. Estudiante de doctorado en Ciencias con Orientación en Probabilidad y Estadística.

<sup>b</sup>Investigador 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 t3rmino de error se asume que tiene media cero y varianza constante  $\sigma^2$ . Una vez se especifica el modelo, la siguiente etapa consiste en determinar en qu3 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 est3n regidas por alg3n criterio de optimalidad (con significado estadístico). Hay varios ejemplos pr3cticos que han hecho uso de los diseños 3ptimos (véase Atkinson (1996)) y existe un gran n3mero 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 3ptimo y sus propiedades; y posteriormente realiz3 muchos trabajos en el 3rea (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 3ptimos.

Este trabajo tiene como objetivo presentar los conceptos b3sicos de los diseños 3ptimos 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 3ptimos locales mencionados en la secci3n 3.1.

Este art́culo se divide en cuatro secciones. En la siguiente secci3n se dar3n los aspectos sobresalientes de los diseños 3ptimos 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 3ltima secci3n se construyen diseños 3ptimos 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 3ptimos para modelos lineales

Para los modelos lineales se considera que la relaci3n entre las  $N$ -observaciones  $Y_i$  y  $x_i$  est3 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  $\chi$ , rango de regresi3n,  $\chi \subseteq \mathbb{R}^k$ ,  $\theta \in \mathbb{R}^m$  es un vector de  $m$ -parámetros desconocidos,  $\epsilon$  es una variable aleatoria con media cero y varianza constante  $\sigma^2$  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  $\chi$  que incluye los conjuntos unitarios; tal que  $\xi$  tiene soporte finito. El soporte de  $\xi$  es  $\text{Supp}(\xi) = [x_1, \dots, x_n]$ ,  $n$ : número de puntos de soporte de  $\xi$ , 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  $\xi$  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  $\Xi$ ; 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  $\theta$ ;  $C_K$  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  $\phi$ . 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  $\theta$ . Es decir, la estimación del subsistema  $K^T\theta$ , donde  $K \in \mathbb{R}^{m \times q}$  es conocida y  $r(K) = q$ .

- Sea  $\xi$  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  $\theta$ , 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  $\phi$ .

Sea  $\phi$  un funcional de valor real,  $\phi : NND(q) \rightarrow \mathbb{R}$ .  $\phi$  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  $\phi$ , 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  $\xi^*$  que sea factible y que maximice, sobre todos los dise1os  $\xi$  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  $\phi$  y  $C_K$ , principalmente la semicontinuidad superior y la compacidad de  $\chi$ , el m3ximo anterior se alcanza para alg3n dise1o  $\xi$ .

- $c$ -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  $\xi^*$  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  $\phi_p$ -óptimos.

- $\phi_p$  es función de información para  $p \in [-\infty, 1]$ .
- Si un diseño  $\xi$  maximiza el criterio anterior, se dice que el diseño es  $\phi_p$ -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  $\theta$ , 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  $\phi$ -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 –  $\xi$  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  $\xi$  minimiza la varianza más grande posible sobre  $\chi$ , rango de regresión.

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

Sea  $\chi \subseteq \mathbb{R}^k$  con  $m$ -vectores linealmente independientes. Un diseño  $\xi$  con matriz de momentos  $M(\xi)$ , definida positiva, es  $D$ -óptimo si y sólo si  $\xi$  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  $\phi_p$  ( $p \in [-\infty, 1]$ ), se enuncia el siguiente teorema de equivalencia, da condiciones necesarias y suficientes para garantizar que un diseño dado es  $\phi_p$ -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  $\xi$  factible para  $K^T \theta$  y matriz de información  $C = C_K(M(\xi))$ . Entonces:*

- $\xi$  es  $\phi_p$ -óptimo 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  $\widetilde{M} \in \mathcal{M}$   $\phi_p$ -óptima para  $K^T\theta$  en  $\mathcal{M}$ .

- Si  $0 < M(\xi) \in \mathcal{M}$ , entonces  $\xi$  es  $\phi_p$ -óptimo para  $\theta$  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 condición 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 condición 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 versión del teorema de equivalencia para  $E$ -optimalidad ( $p = -\infty$ ) (véase Pukelsheim 1993). Para  $p = -1$  ( $A$ -optimalidad),  $M > 0$  y  $C = (K^TM^{-1}K)^{-1}$ , la condición a verificar será:

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

### 3. Diseños óptimos 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 diseño compacto,  $\chi \subseteq \mathbb{R}^k$ , dotado de una  $\sigma$ -álgebra,  $\mathcal{B}$ , (Borelianos en  $\chi$ , agregándole los conjuntos unitarios),  $\theta \in \Theta \subseteq \mathbb{R}^m$ , los errores con media cero y varianza constante y  $\eta(x, \theta)$  es una función no lineal en  $\theta$ .

En el modelo 4, dado un diseño  $\xi$  definido en  $\mathcal{B}$ , se sabe que el estimador de mínimos cuadrados para  $\theta$ , bajo ciertas condiciones de regularidad, es asintóticamente insesgado y su matriz de varianzas-covarianzas asintótica 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 análisis de  $M(\xi, \theta)$ . En la literatura a  $M$  se le conoce como matriz de información, y juega el papel de la matriz de momentos del modelo lineal, si se considerara el modelo linealizado.

La dependencia de  $M$  de  $\theta$  hace que la búsqueda de diseños óptimos dependa de este parámetro. En forma análoga al caso lineal, se cuantifica la magnitud de la información suministrada por  $M(\xi, \theta)$  a partir de funcionales de esta, y consecuentemente la maximización de alguna función de información  $\phi$ , de valor real. Para la construcción de los diseños óptimos 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  $\theta$ ,  $\theta_0$ , que esté cercano al valor verdadero del parámetro, luego utilizar la aproximación lineal de Taylor para  $\eta(x, \theta)$  alrededor de  $\theta_0$  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 $\pi$ -enfoque Bayesiano

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

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

En general, un diseño es  $\phi$ -óptimo Bayesiano con respecto a la distribución a priori  $\pi$ , abreviado por  $\phi_\pi$ -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  $D_\pi$  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  $A_\pi$ -optimalidad al calcular la esperanza, con respecto a  $\theta$ , 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  $\theta$ .

## 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  $\phi_p$  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  $\sigma^2$ , y los  $l$  vectores de regresión  $\{x_1, x_2, \dots, x_l\}$  linealmente independientes y *conocidos*.

El objetivo es encontrar un diseño experimental  $\xi$  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:

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

$\Delta_w^-$  inversa generalizada para  $\Delta_w$ , 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  $\phi_p$ -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  $\theta$ , 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  $\sigma^2$ .

**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  $\mu_2$ . Aplicando el teorema 3 se obtienen los pesos  $\phi_p$ -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 $\mu$	-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 $\mu_2$	-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  $\theta$  (tomando  $K = I$ ).

En efecto, bastará con mostrar que el diseño  $\xi$  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  $\xi$  es  $D$ -3ptimo para estimar el vector de parámetros  $\theta$ .

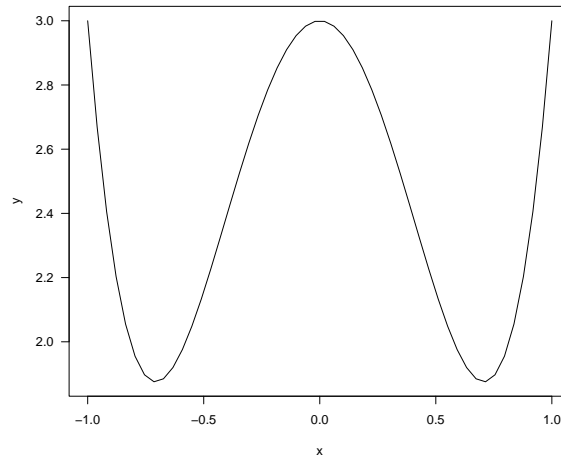


FIGURA 1: Gráfico de la funci3n  $d(x, \xi)$ .

Suponga que el interés del investigador est́ en estimar la diferencia entre el coeficiente de la potencia cuadrática y la lineal. En este caso el sistema de interés 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 número 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 deberán tomar en  $x = 0$ , y el resto se reparte equitativamente en los otros dos puntos, mientras que con  $D$ -optimalidad el mismo número de observaciones se deberá 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 $\theta$	-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  $\theta_0$  es una buena asignación para  $\theta$ , 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  $\xi_0$  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  $\theta_0$ , y puede llegar a ser ineficiente si  $\theta_0$  está muy lejos del valor verdadero  $\theta$ . Otra forma de hallar un diseño óptimo es a partir de un enfoque Bayesiano, donde se incorpora el conocimiento acerca de  $\theta$  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  $D_\pi$ –óptimos para estimar  $\theta$ , 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  $D_\pi$ –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  $\theta$ .

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  $\theta$ , estimar tres cantidades que ayudan al estudio de la cinética del medicamento en un individuo. Estas cantidades son:

- El área bajo la curva (AUC):  $g_1(\theta) = \int_0^\infty \eta(\theta, x) dx = \frac{1}{\theta_2}$ .
- Tiempo para la concentración máxima:  $g_2(\theta) = x_{\max} = \frac{\log \theta_1 - \log \theta_2}{\theta_1 - \theta_2}$ .
- 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  $\theta_0$ . 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  $\theta$ :

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

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  $\pi$ , 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  $\pi$  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 $\pi$	[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.

# Diseños experimentales óptimos en modelos de compartimientos <sup>1</sup>

Víctor Ignacio López Ríos <sup>2</sup>

*Estudiante de Doctorado en Probabilidad y Estadística- Centro de Investigación en Matemáticas- CIMAT. Profesor Escuela de Estadística, Universidad Nacional de Colombia*

Rogelio Ramos Quiroga <sup>3</sup>

*Investigador Titular A, Centro de Investigación en Matemáticas (CIMAT)*

## 1. Introducción

El propósito de este trabajo es la construcción de diseños óptimos para la estimación de los parámetros en un modelo de cuatro compartimientos con tasas de transferencia reversibles; además de diferentes funciones no lineales de éstos. Los modelos de compartimientos son usados para el análisis de un sistema dividido en un número finito de componentes, llamados compartimientos. Estos modelos son de gran utilidad en farmacocinética (rama de la farmacología que se ocupa de la liberación, absorción, distribución, metabolismo y excreción de los medicamentos desde el organismo), y su interés está en el estudio de las curvas concentración-tiempo de un medicamento, a partir de estos modelos.

Estamos interesados en hallar tiempos de muestreo “óptimos” con el fin de estimar, entre otras cantidades, el área bajo la curva concentración-tiempo (AUC) (importante en la deducción de parámetros farmacocinéticos como el aclaramiento -volumen de sangre o plasma depurado de medicamento en unidad de tiempo-, el volumen de distribución - forma en la que un medicamento podría llegar a penetrar en tejidos y líquidos orgánicos- Clark y Smith 1989), el tiempo para la concentración máxima ( $t_{\text{máx}}$ ) y la concentración máxima ( $C_{\text{máx}}$ ).

El modelo de estudio en este trabajo, se representa mediante el siguiente esquema:



---

<sup>1</sup>Trabajo realizado con apoyos de Centro de Investigación en Matemáticas (CIMAT), México, Secretaría de Relaciones Exteriores de México, y Universidad Nacional de Colombia, Sede Medellín.

<sup>2</sup>[viloquez@ciamat.mx](mailto:viloquez@ciamat.mx)

<sup>3</sup>[rrososq@ciamat.mx](mailto:rrososq@ciamat.mx)

siendo los  $\theta_i$  las tasas de transferencia del medicamento de un compartimiento a otro, se supone que la tasa del compartimiento receptor es proporcional a la concentración en el compartimiento fuente. En  $B$  y  $C$  se tienen tasas reversibles. Varios modelos similares, con tasas reversibles, fueron considerados en Allen, D. M (1983), Waterhouse, T. H. (2005), y sin tasa reversible está el trabajo de Atkinson et. al (1993).

El esquema 1 se expresa por medio del siguiente sistema de ecuaciones diferenciales:

$$\begin{aligned} \frac{d\eta_A(t, \Theta)}{dt} &= -\theta_5\eta_A(t, \Theta) & \frac{d\eta_B(t, \Theta)}{dt} &= \theta_1\eta_D(t, \Theta) - \theta_2\eta_B(t, \Theta) + \theta_3\eta_C(t, \Theta) \\ \frac{d\eta_D(t, \Theta)}{dt} &= \theta_5\eta_A(t, \Theta) - \theta_1\eta_D(t, \Theta) & \frac{d\eta_C(t, \Theta)}{dt} &= \theta_2\eta_B(t, \Theta) - (\theta_3 + \theta_4)\eta_C(t, \Theta), \end{aligned}$$

siendo  $\Theta^T = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$  y  $\eta_R$ : concentración del medicamento en el compartimiento  $R$ . Como el objetivo es estimar las cantidades de interés en el compartimiento  $C$ , se presenta sólo la solución para este compartimiento:

$$\eta_C(t, \Theta) = g_1(\Theta) \left[ g_2(\Theta)e^{-\theta_1 t} + g_3(\Theta)e^{-\theta_5 t} + g_4(\Theta)e^{k_2^{(1)} t} + g_5(\Theta)e^{k_2^{(2)} t} \right] \quad (2)$$

$\gamma = \frac{\theta_1 C_0 \theta_5}{\theta_5 - \theta_1}$  y  $(\theta_2 + \theta_3 + \theta_4)^2 - 4\theta_2\theta_4 > 0$  y  $k_2^{(1)}, k_2^{(2)}$  soluciones a:  $k_2^2 + k_2(\theta_2 + \theta_3 + \theta_4) + \theta_2\theta_4 = 0$ .

Y las funciones  $g_i(\Theta)$  están dadas por:

$$\begin{aligned} g_1(\Theta) &= \frac{\theta_1 C_0 \theta_5 (k_2^{(2)} + \theta_2) (k_2^{(1)} + \theta_2)}{\theta_3 (\theta_5 - \theta_1) (k_2^{(1)} - k_2^{(2)})}, & g_2(\Theta) &= \frac{k_2^{(2)} - k_2^{(1)}}{(k_2^{(1)} + \theta_1) (k_2^{(2)} + \theta_1)}, & g_3(\Theta) &= \frac{k_2^{(1)} - k_2^{(2)}}{(k_2^{(1)} + \theta_5) (k_2^{(2)} + \theta_5)}, \\ g_4(\Theta) &= \frac{\theta_1 - \theta_5}{(k_2^{(1)} + \theta_1) (k_2^{(1)} + \theta_5)} & \text{y} & & g_5(\Theta) &= \frac{\theta_5 - \theta_1}{(k_2^{(2)} + \theta_1) (k_2^{(2)} + \theta_5)}. \end{aligned}$$

Se denotarán por  $H_i(\Theta)$ ,  $i = 1, 2, 3, 4, 5$ , las funciones asociadas al área bajo la curva de  $\eta_C(t, \Theta)$ , (AUC), tiempo para la concentración máxima,  $t_{\max}$ , concentración máxima ( $\eta(t_{\max}, \Theta)$ ), y además, la diferencia entre las tasas reversibles,  $(\theta_2 - \theta_3)$  y la suma de las todas las tasas,  $(\sum \theta_j)$ .

## 2. Preliminares y Definiciones

Considere un experimento con  $N$  observaciones independientes  $Y_i$ ,  $i = 1, 2, \dots, N$ , en las condiciones experimentales  $t_i$ , modelado por

$$Y_i = \eta(t_i, \Theta) + \epsilon_i, \quad i = 1, 2, \dots, N, \quad t_i > 0, \Theta \in \mathbb{R}^p, \quad (3)$$

donde los errores  $\epsilon_i$  se suponen independientes e idénticamente distribuidos con media cero y varianza común,  $\sigma^2$ . Un diseño aproximado

$$\xi = \begin{pmatrix} t_1 & \dots & t_n \\ w_1 & \dots & w_n \end{pmatrix}$$

es una medida de probabilidad con soporte finito, es decir, las observaciones son tomadas en los puntos de soporte  $t_i$  proporcional al peso  $w_i$ . Para el modelo descrito en 3, y bajo algunas suposiciones de regularidad, el estimador de mínimos cuadrados es asintóticamente insesgado con matriz de covarianza asintótica proporcional a la inversa de:

$$M(\xi, \Theta) = \sum_{i=1}^n w_i f(t_i, \Theta) f^T(t_i, \Theta) = \int f(t, \Theta) f^T(t, \Theta) d\xi(t),$$

donde  $f(t, \Theta) = \frac{\eta(t, \Theta)}{\partial \Theta}$ . En la literatura de diseños, la matriz  $M(\xi, \Theta)$  se conoce como la matriz de información del diseño  $\xi$  y un criterio de optimalidad,  $\phi$ , maximiza algún funcional de valor real (con algún significado estadístico) de la matriz de información sobre la clase de todos los diseños aproximados (ver Pukelsheim 1993). Ejemplos de tales criterios son:  $\phi(\xi) = \ln |M(\xi, \Theta)|^{1/p}$  ( $D$ -optimalidad), y  $\phi(\xi) = \left\{ \text{Tr} \left[ AM(\xi, \Theta)^{-1} \right] \right\}^{-1}$ , con  $A$  una matriz definida positiva ( $L$ -optimalidad).

El último criterio es apropiado para la estimación de combinaciones lineales de los parámetros,  $K^T \Theta$ , ya que equivale a minimizar el promedio de las varianzas de las combinaciones lineales estimadas, con  $A = K^T K$ . Un diseño  $\xi$  que maximiza  $\phi$  se dice que es  $\phi$ -óptimo para estimar  $\Theta$ . En este trabajo, para la estimación conjunta de las cantidades de interés,  $H_i(\Theta)$ , funciones no lineales de  $\Theta$  ( $i = 1, 2, 3$ ), y lineales ( $i = 4, 5$ ); se va a considerar una versión linealizada, tomando  $A = K^T(\Theta)K(\Theta)$ , donde la columna  $i$  de  $K$  es el gradiente de  $H_i(\Theta)$ . La matriz de información y también, en general,  $K$  dependen de  $\Theta$ , una solución es considerar un valor apriori para  $\Theta$ ,  $\Theta_0$ , y los diseños así obtenidos se denominan locales (Chernoff 1953), además si se tiene una distribución apriori  $\pi$  para  $\Theta$ , se define el criterio  $\phi$  optimal promediado por la apriori  $\pi$ , abreviado por  $\phi_\pi$ , como:  $\phi_\pi(\xi) = E[\phi(\xi, \Theta)] = \int \phi(\xi, \Theta) d\pi(\Theta)$ . Si no hay lugar a confusión, en lo que sigue, se omite la dependencia de  $\Theta$  en  $f$ ,  $K$  y  $M$ . Para la construcción y verificación de que un diseño dado,  $\xi$ , es  $\phi$ -óptimo, en la literatura se tienen varios teoremas de equivalencia (Fedorov y Hackl 1997):

$\xi$  es  $\mathbb{D}$ -óptimo local si  $f^T(t)^T M^{-1}(\xi) f(t) - p \leq 0, \forall t \in [0, \infty)$ ,

$\xi$  es  $\mathbb{L}$ -óptimo local si  $f^T(t) M^{-1}(\xi) K K^T M^{-1}(\xi) f(t) - \text{Tr}(K^T M^{-1}(\xi) K) \leq 0, \forall t \in [0, \infty)$ ,

$\xi$  es  $\mathbb{D}_\pi$ -óptimo si  $\int [f^T(t, \Theta) M^{-1}(\xi, \Theta) f(t, \Theta)] d\pi(\Theta) - p \leq 0 \forall t \in [0, \infty)$ .

$\xi$  es  $\mathbb{L}_\pi$ -óptimo si  $E_\Theta [f^T(t, \Theta) M^{-1}(\xi, \Theta) K^T(\Theta) M^{-1}(\xi, \Theta) f(t, \Theta) - \text{Tr}(K^T(\Theta) M^{-1}(\xi, \Theta) K(\Theta))] \leq 0,$

$\forall t \in [0, \infty)$ , en todos los casos anteriores, con igualdad en los puntos de soporte del diseño óptimo.

$\Theta_0$	$D$ -óptimo local					$L$ -óptimo local				
$\Theta_0^1$	{ 0,73 0,20	1,60 0,20	3,05 0,20	5,45 0,20	10,10 0,20	{ 0,544 0,233	1,471 0,152	2,988 0,171	5,719 0,199	12,046 0,245
$\Theta_0^2$	{ 0,80 0,20	1,92 0,20	3,85 0,20	7,15 0,20	13,30 0,20	{ 0,639 0,215	1,797 0,124	3,775 0,141	7,549 0,185	15,797 0,335
Apriori	$D_\pi$ -óptimo					$L_\pi$ -óptimo				
$\pi$	{ 0,730 0,20	1,600 0,20	3,050 0,20	5,450 0,20	10,196 0,20	{ 0,531 0,473	1,493 0,194	3,077 0,121	5,754 0,096	11,579 0,116

Cuadro 1: Diseños óptimos locales y promediados por una apriori  $\pi$  con valores locales para  $\Theta$  de  $\Theta_0^{1T} = (1,1 \ 2,82 \ 2,17 \ 0,54 \ 1,57)$ ,  $\Theta_0^{2T} = (0,5 \ 2,82 \ 2,17 \ 0,54 \ 1,57)$

### 3. Resultados

En esta sección se presentan los diseños óptimos, locales y promediados por una apriori, obtenidos por los diferentes criterios, para la estimación de las funciones  $H_i(\Theta)$  simultáneamente y también para la estimación de todas las tasas de transferencia. Se hizo un programa en  $R$  para la construcción de los diseños óptimos utilizando los algoritmos propuestos por Fedorov y Hackl (1997). Como ilustración se consideraron dos valores apriori para  $\Theta$  y una distribución apriori uniforme discreta  $\pi$ , para  $\Theta_0^1$  y las 32 combinaciones de los valores de éste, obtenidas al perturbar cada componente en un 10%, es decir, cada  $\theta_i$  tomó dos valores  $\theta_{i0}^1 \pm 0,10\theta_{i0}^1$ , donde el soporte de  $\pi$  es:  $\Omega = \{\Theta : \Theta = \Theta_0^1 \pm 0,10\Theta_0^1, \text{ o } \Theta = \Theta_0^1\}$ . En el cuadro 1, se muestran los diferentes diseños obtenidos, donde los pesos de los diseños  $D$ -óptimos locales y  $D_\pi$ -óptimos son iguales a  $1/5$ , lo cual está acorde con la teoría de diseños óptimos. Cada uno de los diseños proporcionan los tiempos de muestreo óptimos y la frecuencia de la toma de muestras en cada uno de los tiempos. Se observan diferencias de los diferentes diseños obtenidos por cada uno de los criterios. Note que los diseños  $L$ -óptimos locales dan un peso mayor a los puntos extremos del diseño hallado, mientras que el diseño  $L_\pi$ -óptimo da un peso del 47% al primer punto. De la Figura 1 se verifica que todos los diseños son óptimos, uso directo de los teoremas de equivalencia de la sección 2.

### 4. Conclusiones

Se construyeron diseños óptimos para la estimación de funciones no lineales de  $\Theta$ , tanto locales como promediados por una apriori, en modelos de compartimientos útiles en farmacocinética, usando  $L$ -optimalidad. Como trabajo futuro se construirán diseños óptimos que permitan estimar las cantidades de interés y también discriminen entre dos modelos de compartimientos.

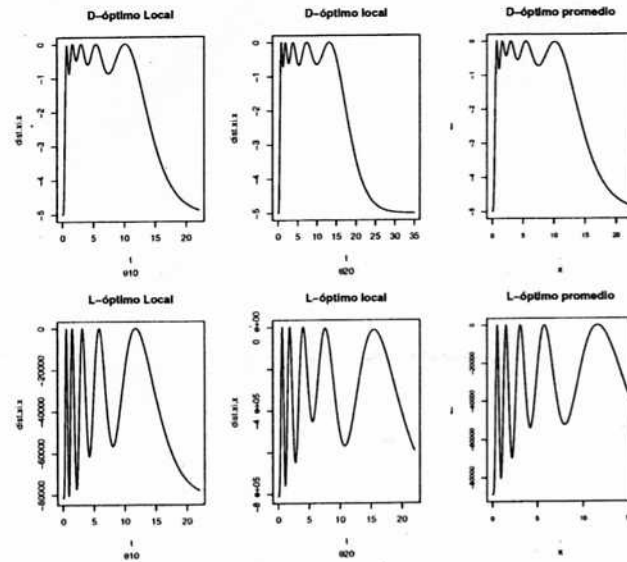


Figura 1: Gráficos para verificar  $\phi$ -optimalidad

## 5. Referencias

- Allen, D. M. (1983). Parameter Estimation for Nonlinear Models with Emphasis on Compartmental Models. *Biometrics*, 39, 629–637.
- Atkinson, A. C., Chaloner, K., Herzberg, A. M. y Juritz, J. (1993). Optimum Experimental Designs for Properties of a Compartmental Model. *Biometrics*, 49 (2), 325–337.
- Chernoff, H. (1953). Locally Optimal Designs for Estimating Parameters. *The Annals of Mathematical Statistics*, 24 (24), 586–602.
- Clark, B. y Smith, D. A. (1989). *An Introduction to Pharmacokinetics*, Oxford.
- Fedorov, V. V. y Hackl (1997). *Model-Oriented Design of Experiments*, New York.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*, New York: Wiley.
- R Development Core Team (2005). R Foundation for Statistical Computing, <http://www.R-project.org>
- Waterhouse, T. H (2005). Optimal Experimental Design for Nonlinear and Generalised Linear Models. *Ph. D. Thesis*. University of Queensland.