



UNIVERSIDAD NACIONAL DE COLOMBIA

Diseños óptimos para estimación y discriminación en modelos de efectos mixtos no lineales

María Eugenia Castañeda López

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2016

Diseños óptimos para estimación y discriminación en modelos de efectos mixtos no lineales

María Eugenia Castañeda López

Tesis presentada como requisito parcial para optar al título de:
Doctora en Ciencias Estadística

Director:
Ph.D. Víctor Ignacio López Ríos

Línea de Investigación:
Diseño de Experimentos
Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2016

Dedicatoria

A Dios y a la Virgen.

A mi esposo Jean y a mi hijo Juan José.

A mis padres Blanca y Hernán.

Agradecimientos

Deseo expresar mis más sinceros agradecimientos a mi director de tesis, Dr. Víctor Ignacio López Ríos, profesor de la Escuela de Estadística de la Universidad Nacional Sede Medellín, por su asesoría y acompañamiento en el desarrollo de este trabajo.

También deseo agradecer a la Universidad de Antioquia por su apoyo en mi proceso de formación académica.

Finalmente, a mi esposo Jean y a mi hijo Juan José, por su amor, por su paciencia y comprensión. A todos mis seres queridos, por su apoyo y confianza. Desde mi fe, infinitas gracias a Dios y a la Virgen.

Resumen

El objetivo de esta Tesis Doctoral es construir diseños óptimos para estimación de parámetros y discriminación entre dos modelos de efectos mixtos no lineales anidados, con diferentes matrices de covarianza intra-individuos. Para resolver este problema se construye un criterio de diseño que combina un criterio para estimación con un criterio para discriminación. Se establece y se demuestra el respectivo teorema de equivalencia. Para el problema de discriminación entre modelos se propone una extensión del criterio de T-optimalidad y se establece el respectivo teorema de equivalencia. Aunque en la literatura existen varias propuestas de métodos para discriminación, la metodología propuesta en esta tesis permite discriminar entre modelos anidados, grupos de individuos con diferentes diseños experimentales y verificación de la optimalidad de un diseño. Por medio de ejemplos se ilustran las metodologías propuestas.

Como resultado de esta Tesis Doctoral se elaboró un artículo de investigación publicado en la revista *Ciencia y Desarrollo* editada por el Centro de Investigaciones y Extensión de la Facultad de Ciencias (CIEC) de la Universidad Pedagógica y Tecnológica de Colombia UPTC, indexada en categoría A2 por Colciencias.

Palabras clave: Modelos de efectos mixtos, Diseños óptimos, D-optimalidad, T-optimalidad, Criterio de diseño compuesto.

Abstract

The objective of this Doctoral Thesis is to find optimal designs for parameter estimation and discrimination between two nested nonlinear mixed effects models which differ in their within-individual covariance matrix. A design criterion that combines a criterion for parameter estimation with a criterion for discrimination is proposed. For this criterion an equivalence theorem is provided. For the problem of discrimination between models a generalization of the T-optimality criterion is proposed and an equivalence theorem is provided. Although there are several proposed methods for discrimination in the literature, the proposed methodology in this thesis can be used for nested models discrimination, for groups of subjects following different designs and for checking the optimality of a proposed design. The application of the proposed methodologies are illustrated with examples.

As result of this Doctoral Thesis one research paper has been elaborated and published on the journal *Ciencia y Desarrollo* published by the Center for Research and Extension of the

Faculty of Sciences (CIEC) of the Pedagogical and Technological University of Colombia UPTC, indexed A2 category by Colciencias.

Keywords: Mixed effects model, Optimal design, D-optimality, T-optimality, Compound design criterion.

Contenido

Agradecimientos	vii
Resumen	viii
1. Introducción	2
2. Teoría de diseños óptimos	6
2.1. Modelo	6
2.2. Estimación	7
2.3. Diseño experimental	8
2.4. Criterios de optimalidad	10
2.5. Teorema de Equivalencia	11
2.6. Diseños óptimos para discriminar entre modelos	12
2.7. Eficiencia de un diseño	15
3. Diseños óptimos para discriminación entre modelos marginales mixtos tipo I	16
3.1. Modelo	17
3.2. Diseño poblacional	20
3.3. Discriminación entre modelos marginales mixtos	22
3.4. Criterio propuesto para dicriminación entre dos modelos intra-individuales	23
3.4.1. Justificación del criterio	24
3.4.2. Diseños T_W^I -óptimos aproximados	26
3.5. Teorema de Equivalencia	27
3.6. Ejemplos	31
3.6.1. Modelo polinomial de decrecimiento exponencial	31
3.6.2. Modelo de dos compartimientos	38
4. Diseños óptimos para discriminación entre modelos marginales mixtos tipo II	43
4.1. Modelo y un criterio de optimalidad	43
4.2. Teorema de equivalencia	46
4.3. Discriminación en modelos no lineales de efectos mixtos	47
4.3.1. Modelo no lineal de efectos mixtos	47
4.3.2. Modelo aproximado lineal	48
4.4. Un Ejemplo	49

5. Diseños óptimos para estimación y discriminación en modelos mixtos	53
5.1. Descripción del problema de estimación y discriminación	53
5.2. Matriz de Información de Fisher	54
5.3. Diseños óptimos para estimación	57
5.4. Diseños para estimación y discriminación	58
5.5. Un ejemplo	61
6. Conclusiones y recomendaciones	68
6.1. Conclusiones	68
6.2. Recomendaciones	69
A. Anexo: Conjuntos convexos y funciones convexas	70
A.1. Conjuntos convexos	70
A.2. Funciones convexas	71
A.3. Diferenciabilidad	71
B. Anexo: Código fuente en R para el cálculo de diseños T_W^I – óptimos	74
Bibliografía	78

1. Introducción

En el campo de la investigación y en muchas áreas de aplicación es una práctica común hacer o realizar experimentos con el fin de resolver un problema o comprobar una idea. El éxito de este experimento radica en gran parte en su planificación, la cual depende entre otros aspectos, del diseño experimental. Un experimento mal diseñado puede conducir a información imprecisa o poco confiable acerca del fenómeno de estudio, y posiblemente el experimento se deba realizar nuevamente implicando costos adicionales. Por esta razón, es importante contar con una metodología estadística que permita determinar aquellos factores que definen el diseño, de forma tal que garanticen una buena respuesta a los interrogantes planteados por el experimentador. La teoría de diseños óptimos resuelve este tipo de problemas, proporcionando las herramientas para la elección de las mejores condiciones experimentales en las cuales se deben realizar las observaciones del proceso o fenómeno de estudio, en el sentido de que éstas maximizan o minimizan algún criterio con significado estadístico.

El problema de diseños óptimos ha sido considerado por más de cien años a través de numerosas contribuciones teóricas y aplicaciones en varias ciencias experimentales. Los primeros resultados de la teoría de diseños óptimos se establecieron con el propósito de construir diseños experimentales que maximizaran la precisión en la estimación de los parámetros del modelo de regresión lineal. Entre ellos, se encuentran, la primera contribución realizada por Smith (1918), quien consideró el problema de diseños para modelos de polinomios hasta de grado seis, llamados posteriormente diseños G-óptimos en Kiefer y Wolfowitz (1960), y el trabajo realizado por Wald (1943) quien introdujo el criterio de D-optimalidad, uno de los criterios más utilizados en la teoría de diseños óptimos. Una de las contribuciones más importantes a la teoría de diseños óptimos es el desarrollo de la teoría de diseños aproximados a través de la formulación de los teoremas de equivalencia. El primero y más utilizado es el Teorema de Equivalencia de Kiefer y Wolfowitz (1960) en el que se establece la equivalencia de los criterios de D-optimalidad y G-optimalidad, además de proporcionar condiciones que permiten verificar la optimalidad de un diseño. El teorema fue generalizado a otros criterios de optimalidad por Whittle (1973) en el Teorema General de Equivalencia. Los libros de Atkinson y Donev (1992), Fedorov y Hackl (1997) y Pukelsheim (1993) conforman un buen compendio de los resultados más importantes de los diseños óptimos.

Un elemento clave en la aplicación de la teoría de diseños óptimos es la definición de un criterio de optimalidad a través del cual se formula el problema de diseño experimental óptimo

como un problema de optimización y cuya elección depende de los objetivos que se buscan al realizar el experimento. Por ejemplo, varios criterios de optimalidad han sido sugeridos para el problema de estimación óptima de los parámetros del modelo, los más importantes y utilizados son aquellos definidos como funciones de la matriz de información de Fisher, como el criterio de D-optimalidad. Si el objetivo es seleccionar el modelo más adecuado entre dos o más modelos competitivos que describen el mismo fenómeno de estudio, se puede utilizar el criterio de T-optimalidad introducido por Atkinson y Fedorov (1975). También se han introducido criterios con múltiples objetivos, como por ejemplo, los criterios compuestos que combinan dos o más criterios de optimalidad, ver por ejemplo Atkinson y Donev (1992). Estos criterios han sido ampliamente desarrollados para modelos de regresión tanto lineales como no lineales.

Una clase de modelos particularmente útiles en estudios longitudinales en los cuales el número de observaciones por individuo es limitado, son los modelos no lineales de efectos mixtos. En estos modelos se distinguen dos clases de variación: la variación aleatoria entre las observaciones dentro de un individuo (intra-individual) y la variación aleatoria entre los individuos (inter-individuos) (ver Davidian y Giltinan, 1995; Demidenko 2004). Esta separación de la variabilidad permite la estimación de las características de la población a partir de un grupo de individuos con muestras escasas sin necesidad de la estimación individual de los parámetros. Dependiendo de la naturaleza de los datos, diferentes estructuras se pueden utilizar para describir la variabilidad intra-individual, por ejemplo, en estudios poblacionales se puede suponer errores heterocedásticos y/o correlacionados.

La aplicación de estos modelos en áreas como la farmacocinética poblacional ha motivado el desarrollo de metodologías para la construcción de diseños óptimos, especialmente en lo que concierne al problema de estimación de parámetros. Bajo un modelo de efectos mixtos, un diseño poblacional está definido por el número de individuos a estudiar y los diseños asociados con cada individuo (número de muestras y tiempos de medición individuales) (Mentré *et al.*, 1997). El problema de construir diseños óptimos poblacionales ha sido considerado por varios autores. Mentré *et al.* (1997) derivaron una expresión de la matriz de información de Fisher para la construcción de diseños D-óptimos usando la expansión en series de Taylor de primer orden del modelo. Estos autores encontraron diseños D-óptimos en un espacio finito de diseños individuales. Retout *et al.* (2003a) desarrollaron una expresión para la matriz de información en el caso de modelos con covariables y errores heteroscedásticos, bajo el mismo enfoque de Mentré *et al.* (1997). Mielke (2011) sugirió otra aproximación para la matriz de información y estudió el efecto de diferentes expresiones en el diseño de experimentos. Stroud *et al.* (2001) y Han y Chaloner (2004) sugirieron un enfoque alternativo para encontrar diseños D-óptimos con base en métodos Bayesianos. Otros enfoques basados en la construcción de un diseño paramétrico fueron desarrollados por Wang (2007) y Dette *et al.* (2010). Los diseños D-óptimos también han sido estudiados bajo diferentes clases de modelos

mixtos entre ellos modelos con intercepto aleatorio (Debushe, 2004), modelos con coeficientes aleatorios (Schmelter, 2007b; Nie, 2007), modelos con observaciones correlacionadas (Dette *et al.*, 2010; Nyberg *et al.*, 2012), y modelos con covariables (Schmelter 2007a; Bogacka *et al.*, 2015). Para la construcción numérica de los diseños se han propuesto diferentes técnicas como por ejemplo técnicas basadas en el algoritmo Simplex (Retout y Mentré, 2003b), el algoritmo de Fedorov-Wynn (Mentré *et al.*, 1997; Retout *et al.*, 2002; Retout *et al.*, 2007) y algoritmos multiplicativos (Holland-Letz *et al.*, 2012).

En contraste con la estimación de parámetros, no hay muchos métodos que aborden el problema de discriminación entre modelos. Entre las contribuciones realizadas en esta dirección se pueden citar los trabajos de Waterhouse *et al.* (2005) quienes propusieron el criterio producto para estimación y discriminación simultáneamente. Sin embargo, para modelos de regresión no lineales, estos diseños pueden ser menos eficientes para discriminar en comparación con los diseños T-óptimos. Vajjah y Duffull (2012) propusieron una metodología para encontrar diseños T-óptimos robustos que no depende de la selección a priori del modelo correcto. La construcción de estos diseños se basa sólo en los modelos de efectos fijos sin error residual. Kuczewski *et al.* (2008) propusieron una extensión del criterio de T-optimalidad para modelos heteroscedásticos con múltiples respuestas en el caso particular donde todos los individuos son observados bajo las mismas condiciones experimentales.

El problema de diseños con múltiples objetivos, en especial estimación de parámetros y discriminación en modelos, también ha sido objeto de estudio en el caso de modelos de efectos mixtos. Varios de los enfoques propuestos se desarrollan aplicando las metodologías para modelos de regresión. Por ejemplo, Jones y Wang (1999) sugirieron el uso de diseños compuestos combinando T-optimalidad con el criterio D-optimalidad para modelos de efectos mixtos, Waterhouse *et al.* (2005) sugirieron el uso del criterio producto y Vajjah y Duffull (2012) propusieron diseños condicionales combinando D-optimalidad para modelos mixtos y T-optimalidad para las componentes fijas.

El objetivo de este trabajo es construir diseños óptimos para estimación de parámetros y discriminación entre dos modelos no lineales de efectos mixtos anidados, con diferentes matrices de covarianza intra-individuos. Para resolver este problema se construye un criterio de optimalidad que representa un balance entre la estimación de parámetros y la discriminación entre los modelos. Aplicando la teoría de diseños óptimos aproximados se establece y se demuestra el respectivo teorema de equivalencia.

El problema de discriminación para esta clase de modelos anidados, de acuerdo con la revisión bibliográfica realizada, no ha sido considerado y aunque en la literatura existen varias propuestas de métodos para discriminación, como los descritos anteriormente, la metodología propuesta en esta tesis permite la discriminación entre modelos anidados, la posibilidad de

grupos de individuos con diferentes diseños experimentales y la verificación de la optimalidad de un diseño. En este trabajo se propone una extensión del criterio de T-optimalidad para este problema de discriminación. Aplicando la teoría de diseños aproximados se establece y se demuestra el respectivo teorema de equivalencia.

El trabajo se desarrolla a través de los siguientes capítulos. En el capítulo 2 se presentan algunos conceptos y resultados de la teoría de diseños óptimos para modelos de regresión. Aunque el contenido se centra en la estimación de parámetros, éste es fundamental para el desarrollo teórico de los capítulos posteriores.

En los capítulos 3 y 4 se presenta uno de los resultados principales de este trabajo. Específicamente, en el capítulo 3 se considera la discriminación entre una clase particular de modelos mixtos. Se enuncia y se demuestra el respectivo teorema de equivalencia. Para ilustrar la metodología propuesta se presentan dos ejemplos de aplicación: el primero para modelos polinomiales de decrecimiento exponencial y el segundo para modelos de compartimientos. En el capítulo 4 los resultados previos desarrollados en el capítulo 3 se extienden a una clase más general de modelos mixtos. En este caso también se establece el respectivo teorema de equivalencia y el criterio propuesto se ilustra con un ejemplo en modelos de un compartimiento.

En el capítulo 5 se presenta el segundo resultado importante de este trabajo. Se construye un criterio compuesto para los objetivos simultáneos: estimación y discriminación. Se establece el respectivo teorema de equivalencia y se ilustra la aplicación del criterio con un ejemplo.

Finalmente, en el capítulo 6 se presentan las conclusiones del trabajo y algunas líneas de investigación.

2. Teoría de diseños óptimos

En este capítulo se presentan algunos conceptos y resultados de la teoría de diseños óptimos para modelos de regresión sin componentes aleatorias. Aunque el tema abordado se centra principalmente en la estimación de los parámetros del modelo, los conceptos y propiedades que se exponen son fundamentales para la comprensión y desarrollo de los capítulos siguientes. La mayor parte del contenido de este capítulo ha sido tomado de los libros de Silvey (1980), Fedorov y Hackl (1997), y Atkinson y Donev (1992). Como bibliografía complementaria de la teoría de diseños óptimos se recomienda el texto de Pukelsheim (1993) para un tratamiento matemático más riguroso. Para una mayor comprensión de estos resultados así como de los teoremas y demostraciones que se derivan en los capítulos posteriores, en el Anexo A se presentan algunas propiedades importantes de la teoría de conjuntos convexos y funciones convexas.

A continuación se presenta una descripción resumida de la teoría de diseños óptimos en modelos de regresión

2.1. Modelo

Suponga que las observaciones de un experimento se pueden describir mediante el siguiente modelo

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \epsilon_i, \quad i = 1, \dots, n \quad (2-1)$$

donde \mathbf{x}_i es un vector k -dimensional de variables explicativas o regresoras seleccionadas libremente por el experimentador, $\boldsymbol{\beta}$ es un vector p -dimensional de parámetros desconocidos, ϵ_i es el término del error y $f(\mathbf{x}, \boldsymbol{\beta})$ es una función conocida la cual puede ser lineal o no lineal en el vector de parámetros $\boldsymbol{\beta}$. Se asume que el vector \mathbf{x}_i pertenece a un conjunto compacto \mathcal{X} de \mathbb{R}^k y que la función f es continua en \mathcal{X} . El conjunto \mathcal{X} contiene todos los posibles puntos donde se pueden obtener las observaciones y es llamado espacio o región de diseño. Además se supone que los errores son independientes e idénticamente distribuidos con media cero y varianza constante σ^2 .

Para modelos lineales la función de respuesta está dada por

$$f(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{g}(\mathbf{x}_i)^T \boldsymbol{\beta}, \quad (2-2)$$

donde $\mathbf{g}(\mathbf{x}_i) = (g_1(\mathbf{x}_i), \dots, g_p(\mathbf{x}_i))^T$ es un vector de funciones de regresión continuas, de valor real y linealmente independientes. El modelo se puede expresar en forma matricial como

$$\mathbf{y} = \mathbf{G}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2-3)$$

donde $\mathbf{y} = (y_1, \dots, y_n)^T$ es el vector de observaciones, $\mathbf{G} = (\mathbf{g}(\mathbf{x}_1), \dots, \mathbf{g}(\mathbf{x}_n))^T$ denota la matriz de diseño de orden $(n \times p)$ y $\boldsymbol{\epsilon}$ es el vector de errores. Bajo las formulaciones anteriores se tiene que $E(\mathbf{y}) = \mathbf{G}\boldsymbol{\beta}$ y $Cov(\mathbf{y}) = \sigma^2 \mathbf{I}_n$ donde \mathbf{I}_n es la matriz identidad de orden $(n \times n)$.

2.2. Estimación

Una vez definido el modelo, el siguiente paso es ajustar el modelo con los datos experimentales, es decir, determinar los valores de los parámetros que mejor describen los datos. Suponga que se tiene interés en estimar el vector de parámetros $\boldsymbol{\beta}$. Entonces aplicando el método de mínimos cuadrados se obtiene que:

$$\hat{\boldsymbol{\beta}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{y}. \quad (2-4)$$

Para evaluar la calidad del estimador se puede examinar su matriz de covarianzas, la cual bajo los supuestos del modelo está dada por:

$$Cov(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{G}^T \mathbf{G})^{-1}. \quad (2-5)$$

La matriz $\mathbf{G}^T \mathbf{G}$ es llamada la matriz de información para el vector de parámetros $\boldsymbol{\beta}$. Nótese que esta matriz está determinada por las condiciones experimentales y no depende de las observaciones. Así, la precisión en la estimación dependerá considerablemente del diseño experimental.

En el caso de modelos no lineales, generalmente, no es posible hallar una expresión cerrada para $\hat{\boldsymbol{\beta}}$ y excepto en casos especiales, es difícil encontrar una expresión analítica para la matriz de varianzas y covarianzas $Cov(\hat{\boldsymbol{\beta}})$. Para obtener una expresión aproximada de esta matriz se linealiza el modelo mediante un desarrollo en series de Taylor de la función $f(\mathbf{x}, \boldsymbol{\beta})$ alrededor de un valor local del vector de parámetros, $\boldsymbol{\beta}^0$, siendo $\boldsymbol{\beta}^0$ una buena aproximación para el vector de parámetros.

Esta aproximación es

$$f(\mathbf{x}, \boldsymbol{\beta}) \cong f(\mathbf{x}, \boldsymbol{\beta}^0) + (\boldsymbol{\beta} - \boldsymbol{\beta}^0)^T \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\beta}^0), \quad \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\beta}) = \frac{\partial f(\mathbf{x}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}. \quad (2-6)$$

A partir de esta expresión, se puede demostrar que la matriz de varianzas y covarianzas asintótica de $\hat{\boldsymbol{\beta}}$ es (ver Fedorov y Hackl, 1997, Capítulo 1, p. 18):

$$Cov(\hat{\boldsymbol{\beta}}) \cong \left(\sigma^{-2} \sum_{i=1}^n \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\beta}^0) \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\beta}^0)^T \right)^{-1}. \quad (2-7)$$

Por simplicidad, en lo que sigue se considera el caso de modelos lineales.

2.3. Diseño experimental

A partir de la ecuación (2-5), se observa que al seleccionar variables de diseño \mathbf{x} que “maximizan” funcionales de la matriz $\mathbf{G}^T \mathbf{G}$ se obtienen variables de diseño que “minimizan” funcionales de la matriz de varianzas y covarianzas del estimador. Por lo tanto, si el objetivo del experimento es obtener buenos estimadores de los parámetros o funciones de éstos, los niveles de los \mathbf{x}_i y las frecuencias de medición de la respuesta se deben seleccionar de manera que maximicen algún funcional de la matriz de información. Estos valores de \mathbf{x} y sus respectivas frecuencias es lo que define un diseño óptimo.

A continuación se presenta la definición formal de un diseño experimental y el problema de diseño.

Definición 2.3.1. *Un diseño exacto de n observaciones es un conjunto de n puntos de \mathcal{X} representado por $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$.*

Si hay repeticiones de los puntos \mathbf{x}_i entonces el diseño se puede representar como

$$\xi_n = \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_s \\ w_1 & \cdots & w_s \end{array} \right\}, \quad (2-8)$$

donde $w_i = \frac{r_i}{n}$ y r_i es el número de repeticiones del punto \mathbf{x}_i en la muestra de n observaciones. Este diseño es llamado diseño exacto normalizado. Los puntos \mathbf{x}_i se dice que son puntos de diseño o de soporte y los w_i son llamados pesos. Asociada a este diseño se define la matriz de información normalizada:

$$M(\xi_n) = \sum_{i=1}^s w_i \mathbf{g}(\mathbf{x}_i) \mathbf{g}(\mathbf{x}_i)^T. \quad (2-9)$$

El problema de hallar un diseño experimental óptimo se puede formular como un problema de optimización en el que se busca encontrar la solución ξ_n^* que verifica:

$$\xi_n^* = \arg \max_{\xi_n} \Phi(M(\xi_n)), \quad (2-10)$$

donde Φ es un funcional seleccionado según el objetivo del experimento y es llamado criterio de optimalidad.

Para modelos no lineales, debido a que la matriz de información depende del vector de parámetros desconocido β , el problema de diseño (2-10) se puede resolver, por ejemplo, fijando un valor específico del parámetro. En este caso, se dice que el diseño es óptimo localmente (ver Chernoff, 1953).

Con el fin de simplificar el problema de optimización entera (2-10) y caracterizar la solución optimal, Kiefer (1959) introdujo la teoría de diseños óptimos aproximados, en la cual se considera una definición más general de diseño. A esta clase de diseños se les llama diseños aproximados o continuos.

Definición 2.3.2. *Un diseño aproximado o continuo es cualquier medida de probabilidad ξ definida en la sigma álgebra de Borel de \mathcal{X} con soporte finito.*

El soporte de un diseño aproximado ξ está dado por el conjunto de puntos de \mathcal{X} para los cuales los pesos $w_i = \xi(\mathbf{x}_i)$ son distintos de cero, esto es

$$\text{Supp}(\xi) = \{\mathbf{x}_i \in \mathcal{X} : \xi(\mathbf{x}_i) > 0\}. \quad (2-11)$$

El conjunto de todos los diseños aproximados se denota por $\Xi(\mathcal{X})$.

Para un diseño aproximado ξ , la matriz de información normalizada se define como:

$$M(\xi) = \int_{\mathcal{X}} \mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^T \xi(d\mathbf{x}). \quad (2-12)$$

Si $\mathcal{M} = \{M(\xi) : \xi \in \Xi(\mathcal{X})\}$ es el conjunto de las matrices de información de todos los diseños aproximados en $\Xi(\mathcal{X})$, entonces el problema de diseño descrito previamente se puede formular como:

$$\xi^* = \arg \max_{\xi \in \Xi(\mathcal{X})} \Phi(M(\xi)). \quad (2-13)$$

Los diseños que maximizan $\Phi(M(\xi))$ son llamados Φ -optimales.

Las siguientes son algunas características de las matrices de información para diseños aproximados fundamentales en el problema de diseño experimental (ver Fedorov y Hackl, 1997, Capítulo 2, p. 29).

Teorema 2.3.1. *Sea ξ un diseño aproximado en Ξ . Entonces*

- (i) $M(\xi)$ es una matriz simétrica y definida no negativa.
- (ii) $M(\xi)$ es singular si el soporte de ξ contiene menos puntos que el número de parámetros.
- (iii) \mathcal{M} es un conjunto convexo.
- (iv) Si \mathcal{X} es compacto y g es continua en \mathcal{X} entonces \mathcal{M} es compacto.
- (v) Para cualquier matriz M de \mathcal{M} existe un diseño ξ que no contiene más de $n_0 = p(p+1)/2 + 1$ puntos con pesos diferentes de cero, tal que $M(\xi) = M$. Si M es un punto de frontera de \mathcal{M} entonces $n_0 = p(p+1)/2$.

Finalmente, bajo el supuesto de normalidad de los errores, la matriz de información $M(\xi)$ coincide con la matriz de información de Fisher:

$$M(\xi) = E_{\xi} \left[\frac{\partial}{\partial \boldsymbol{\beta}} \log p(y|\mathbf{x}, \boldsymbol{\beta}) \left(\frac{\partial}{\partial \boldsymbol{\beta}} \log p(y|\mathbf{x}, \boldsymbol{\beta}) \right)^T \right] = E_{\xi} \left[-\frac{\partial^2 \log p(y|\mathbf{x}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right], \quad (2-14)$$

donde $p(y|\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y - g(\mathbf{x})^T \boldsymbol{\beta})^2}{2\sigma^2} \right\}$.

2.4. Criterios de optimalidad

Como se mencionó en la sección anterior la función $\Phi(M)$ en el problema de diseño (2-13) se elige acorde con las necesidades del experimento. Con el fin de establecer resultados teóricos y construir algoritmos de optimización es deseable que estas funciones tengan ciertas características que se describen a continuación.

Sea Φ una función real definida en el conjunto de las matrices definidas no negativas de orden $(p \times p)$ denotado por $NND(p)$. Las siguientes son condiciones generales deseables en un criterio de optimalidad:

- (i) Φ es homogéneamente positiva: $\Phi(\delta M) = \delta \Phi(M)$ donde $\delta > 0$.
- (ii) Φ es una función cóncava.
- (iii) Φ es una función monótona: si $M_1 \geq M_2$ entonces $\Phi(M_1) \geq \Phi(M_2)$. La relación entre las matrices se refiere al orden de Loewner.
- (iv) Existe un diseño ξ tal que $\Phi(M(\xi))$ toma un valor finito.
- (v) Φ es Fréchet diferenciable.

En la literatura de diseños óptimos varios criterios de optimalidad han sido propuestos. Algunos de los más utilizados en diferentes aplicaciones son:

- D-optimalidad: $\Phi(M) = \log \det M$. Es el criterio de optimalidad más conocido. Bajo el supuesto de normalidad de las observaciones el diseño D -óptimo minimiza el volumen del elipsoide de confianza del vector de parámetros β .
- A-optimalidad: $\Phi(M) = -\text{tr}(M^{-1})$. Este criterio minimiza el promedio de las varianzas de los estimadores de los parámetros.
- E-optimalidad: $\Phi(M) = \lambda_{\min}(M)$ donde λ_{\min} es el mínimo valor propio de M . El criterio minimiza el mayor de los ejes del elipsoide de confianza del vector de parámetros β .
- c-optimalidad: $\Phi(M) = -c^T M^{-1} c$. Este criterio es útil cuando el objetivo es estimar de forma eficiente una función lineal del vector de parámetros.
- G-optimalidad: $\Phi(M) = -\max_{\mathbf{x} \in \mathcal{X}} d(\mathbf{x}, \xi)$ donde $d(\mathbf{x}, \xi) = \mathbf{g}(\mathbf{x})^T M^{-1} \mathbf{g}(\mathbf{x})$. El criterio minimiza el valor más grande posible de la varianza en el espacio de diseño.

En la siguiente sección se presenta el Teorema General de Equivalencia uno de los resultados más importantes en la teoría de los diseños óptimos. Este teorema no sólo proporciona condiciones precisas para verificar la Φ -optimalidad de un diseño sino que también juega un papel importante en la construcción de algoritmos para hallar el diseño óptimo.

2.5. Teorema de Equivalencia

La equivalencia entre los diseños D- y G-óptimos fue establecida en el Teorema de Equivalencia de Kiefer y Wolfowitz el cual se enuncia a continuación (ver Fedorov y Hackl, 1997, Capítulo 2, p. 34).

Teorema 2.5.1. (Teorema de Equivalencia de Kiefer y Wolfowitz). *Las siguientes afirmaciones son equivalentes:*

- (i) $\xi^* = \arg \max_{\xi \in \Xi(\mathcal{X})} \det M(\xi)$.
- (ii) $\xi^* = \arg \min_{\xi \in \Xi(\mathcal{X})} \max_{\mathbf{x} \in \mathcal{X}} d(\mathbf{x}, \xi)$.
- (iii) $\max_{\mathbf{x} \in \mathcal{X}} d(\mathbf{x}, \xi^*) = p$

A continuación se presentan algunos resultados importantes que conllevan a la formulación del teorema para otros criterios de optimalidad (ver Silvey, 1980, Capítulo 3, pp. 17-23). Primero se introduce una definición alternativa de la derivada de Fréchet.

Definición 2.5.1. *La derivada de Fréchet de Φ en $M(\xi_1)$ en la dirección de $M(\xi_2)$ está definida por*

$$F_{\Phi}(M(\xi_1), M(\xi_2)) = \lim_{\alpha \rightarrow 0^+} \frac{\Phi((1 - \alpha)M(\xi_1) + \alpha M(\xi_2)) - \Phi(M(\xi_1))}{\alpha}. \quad (2-15)$$

En el espacio de las matrices reales de orden $k \times v$ el producto interno está dado por $\langle A, B \rangle = \text{tr}(A^T B)$. Así, si existe el gradiente de Φ entonces

$$F_{\Phi}(M(\xi_1), M(\xi_2)) = \text{tr}[\nabla \Phi(M(\xi_1))(M(\xi_2) - M(\xi_1))]. \quad (2-16)$$

Teorema 2.5.2. *Sea Φ una función criterio cóncava en \mathcal{M} . Entonces ξ^* es Φ -óptimo si y sólo si*

$$F_{\Phi}(M(\xi^*), M(\xi)) \leq 0, \quad \forall \xi \in \Xi(\mathcal{X}). \quad (2-17)$$

Teorema 2.5.3. *Sea Φ una función criterio cóncava en \mathcal{M} y diferenciable en $M(\xi^*)$. Entonces ξ^* es Φ -óptimo si y sólo si*

$$F_{\Phi}(M(\xi^*), \mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^T) \leq 0, \quad \forall \mathbf{x} \in \mathcal{X}. \quad (2-18)$$

Teorema 2.5.4. *Sea Φ una función criterio y sea \mathcal{M}^+ el subconjunto de \mathcal{M} donde $\Phi(M) > -\infty$. Si Φ es diferenciable en todos los puntos de \mathcal{M}^+ y existe una medida Φ -óptima, entonces ξ^* es Φ -óptimo si y sólo si*

$$\max_{\mathbf{x} \in \mathcal{X}} F_{\Phi}(M(\xi^*), \mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^T) = \min_{\xi \in \Xi(\mathcal{X})} \max_{\mathbf{x} \in \mathcal{X}} F_{\Phi}(M(\xi), \mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^T). \quad (2-19)$$

Corolario 2.5.4.1. Si ξ^* es Φ -óptimo y Φ es diferenciable en $M(\xi^*)$ entonces

$$\max_{\mathbf{x} \in \mathcal{X}} F_{\Phi}(M(\xi^*), \mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^T) = 0 \quad (2-20)$$

y

$$E_{\xi}[F_{\Phi}(M(\xi^*), \mathbf{g}(\tilde{\mathbf{x}})\mathbf{g}(\tilde{\mathbf{x}})^T)] = 0, \quad (2-21)$$

donde $\tilde{\mathbf{x}}$ es un vector aleatorio con distribución ξ^* . Esto puede suceder sólo si

$$F_{\Phi}(M(\xi^*), \mathbf{g}(\tilde{\mathbf{x}})\mathbf{g}(\tilde{\mathbf{x}})^T) = 0$$

con probabilidad uno. Así, si ξ^* es una medida discreta con soporte finito $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(I)}$, entonces

$$F_{\Phi}(M(\xi^*), \mathbf{g}(\mathbf{x}_{(i)})\mathbf{g}(\mathbf{x}_{(i)})^T) = 0, \quad i = 1, \dots, I. \quad (2-22)$$

Finalmente, se enuncia el Teorema General de Equivalencia.

Teorema 2.5.5. (Teorema General de Equivalencia). Las siguientes afirmaciones son equivalentes:

(i) ξ^* es Φ -óptimo.

(ii) ξ^* minimiza $\max_{\eta \in \Xi(\mathcal{X})} F_{\Phi}(M(\xi), M(\eta))$.

(iii) $\max_{\eta \in \Xi(\mathcal{X})} F_{\Phi}(M(\xi^*), M(\eta)) = 0$.

2.6. Diseños óptimos para discriminar entre modelos

En la aplicación de la teoría de diseños óptimos uno de los supuestos básicos es asumir que el modelo utilizado para describir las observaciones de un experimento es el modelo correcto. Sin embargo, en algunas situaciones prácticas, pueden existir dos o más modelos potencialmente capaces de proporcionar buenos ajustes para las observaciones. En el contexto de la teoría de diseños óptimos, el problema de elegir cuál es el mejor modelo se puede resolver mediante la construcción de un diseño que tenga una potencia alta para discriminar entre los modelos. Varios enfoques han sido propuestos para la construcción de diseños con este propósito, como por ejemplo, los desarrollados por Pukelsheim y Rosenberger (1993) y López-Fidalgo *et al.* (2007). Una de las metodologías más utilizadas en la literatura de los diseños óptimos fue propuesta por Atkinson y Fedorov (1975), quienes introdujeron los diseños T-óptimos para discriminación entre dos modelos. El criterio se propuso inicialmente para modelos de regresión lineales y ha sido extendido a otras clases de modelos, como por ejemplo, modelos heterocedásticos (Uciński y Bogacka, 2004), modelos con múltiples respuestas (Uciński y Bogacka, 2005), entre otros.

El problema de diseñar un experimento para discriminar entre dos modelos, básicamente consiste en hallar un diseño para el cual las respuestas esperadas bajo cada modelo estén tan alejadas como sea posible de acuerdo con un criterio apropiado. Esta es una forma de decir que posiblemente algunos puntos de diseño proporcionan más información que otros para discriminar entre los dos modelos. Puntos de diseño en los que la diferencia de las respuestas esperadas es grande son puntos que permiten discriminar entre los modelos. En contraste, puntos donde esta diferencia es pequeña son puntos donde los modelos se confunden. Una dificultad que se presenta bajo este enfoque es que dichas diferencias esperadas son desconocidas pues dependen de los valores desconocidos de los parámetros, es más, dependen del vector de parámetros desconocido del modelo correcto que también es desconocido. Una solución a este problema fue propuesta por Atkinson y Fedorov (1975) a través del criterio de T-optimalidad el cual se describe a continuación.

Criterio de T-optimalidad

Considere el modelo de regresión

$$y = f(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon, \quad (2-23)$$

donde el vector de variables explicativas $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}$, con \mathcal{X} un conjunto compacto conocido. La función de regresión f representa el modelo verdadero que relaciona las variables \mathbf{x} y y , y la cantidad ε denota el error aleatorio el cual se asume que es una variable aleatoria con media cero y varianza σ^2 .

Suponga que la función de respuesta $f(\mathbf{x}, \boldsymbol{\beta})$ coincide con uno de dos modelos paramétricos $f_1(\mathbf{x}, \boldsymbol{\beta}_1)$ ó $f_2(\mathbf{x}, \boldsymbol{\beta}_2)$ donde $\boldsymbol{\beta}_1 \in \Theta_1 \subset \mathbb{R}^{p_1}$ y $\boldsymbol{\beta}_2 \in \Theta_2 \subset \mathbb{R}^{p_2}$, con Θ_1 y Θ_2 conjuntos compactos conocidos. El objetivo es diseñar un experimento de manera que las observaciones obtenidas permitan decidir cuál de los dos modelos es el más adecuado. Suponga que se fija uno de los dos modelos como el modelo correcto con su respectivo vector de parámetros conocido, por ejemplo sea $f(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\beta}) = f_1(\mathbf{x}, \boldsymbol{\beta}_1)$. Entonces un diseño T-óptimo aproximado es un diseño que maximiza la desviación mínima entre el modelo f_1 y la clase de modelos definidos por f_2 , esto es, el diseño ξ^* tal que

$$\xi^* = \arg \max_{\xi} \Delta(\xi), \quad (2-24)$$

donde

$$\Delta(\xi) = \int_{\mathcal{X}} (f(x) - f_2(x, \boldsymbol{\beta}_2^*))^2 \xi(dx), \quad (2-25)$$

con

$$\begin{aligned} \boldsymbol{\beta}_2^* &= \arg \min_{\boldsymbol{\beta}_2 \in \Theta_2} \int_{\mathcal{X}} (f(x) - f_2(x, \boldsymbol{\beta}_2))^2 \xi(dx) \\ &= \arg \min_{\boldsymbol{\beta}_2 \in \Theta_2} \Delta(\xi, \boldsymbol{\beta}_2). \end{aligned} \quad (2-26)$$

El valor β_2^* corresponde a la mejor aproximación del modelo $f_1(x, \beta_1)$ determinada por los modelos de la forma $f_2(x, \beta_2)$ donde $\beta_2 \in \Theta_2$. Así, un buen diseño para discriminar entre f_1 y f_2 proporciona una suma de cuadrados de falta de ajuste $\Delta(\xi)$ grande para el segundo modelo. Nótese que los diseños T-óptimos son óptimos localmente pues el criterio depende del vector de parámetros del modelo que se asume como correcto.

Si las funciones $f_j(\mathbf{x}, \beta_j)$ dependen linealmente del vector de parámetros β_j , $j = 1, 2$ y los errores son normales e independientes, entonces la cantidad $\Delta(\xi)$ es proporcional al parámetro de no centralidad de la distribución χ^2 de la suma de cuadrados residual asociada al segundo modelo. Así, los diseños que maximizan (2-25) proporcionan la prueba F más potente para la falta de ajuste del segundo modelo cuando el primer modelo se asume como el modelo correcto. La letra T se usa para enfatizar la conexión con el contraste de hipótesis de modelos (“Testing models”). Para modelos no lineales las propiedades de la prueba F se reemplazan por propiedades asintóticas.

Asumiendo que las funciones $f(\mathbf{x})$ y $f_2(\mathbf{x}, \beta_2)$ son continuas en \mathcal{X} , Atkinson y Fedorov (1975) formularon el respectivo teorema de equivalencia para el criterio $\Delta(\xi)$, el cual se enuncia a continuación (ver Fedorov y Hackl, 1997, Capítulo 5, p. 94).

Teorema 2.6.1. (Teorema de equivalencia para T-optimalidad). *Una condición necesaria y suficiente para que un diseño ξ^* sea T-óptimo es la existencia de una medida ζ^* tal que:*

$$\varphi'(\mathbf{x}, \xi^*) \leq \Delta'(\xi^*) \quad (2-27)$$

donde

$$\varphi'(\mathbf{x}, \xi) = \varphi'(\mathbf{x}, \xi; \zeta) = \int_{\Theta_2(\xi)} \varphi(\mathbf{x}, \beta_2) \zeta(d\beta_2), \quad (2-28)$$

$$\varphi(\mathbf{x}, \beta_2) = (f(\mathbf{x}) - f_2(\mathbf{x}, \beta_2))^2, \quad (2-29)$$

$$\Theta_2(\xi) = \left\{ \beta_2 : \beta_2 = \beta_2(\xi) = \arg \min_{\beta_2 \in \Theta_2} \Delta(\xi, \beta_2) \right\}, \quad (2-30)$$

$$\int_{\Theta_2(\xi)} \zeta(d\beta_2) = 1 \text{ y } \Delta'(\xi) = \min_{\beta_2 \in \Theta_2} \Delta(\xi, \beta_2). \quad (2-31)$$

Además, la función $\varphi'(\mathbf{x}, \xi^*)$ alcanza su valor máximo en los puntos de soporte del diseño ξ^* .

Si el conjunto $\Theta_2(\xi)$ es unitario entonces

$$\varphi'(\mathbf{x}, \xi^*) = (f(\mathbf{x}) - f_2(\mathbf{x}, \beta_2'))^2 \quad (2-32)$$

donde $\beta_2' = \beta_2'(\xi^*)$, lo cual simplifica la verificación de un diseño T-óptimo. Los diseños para los cuales la solución en (2-30) es única son llamados T-regulares.

Otras propiedades de este criterio se pueden encontrar en Atkinson y Donev (1992).

2.7. Eficiencia de un diseño

Una forma de medir la eficiencia de un diseño ξ con respecto a un diseño óptimo ξ^* bajo el mismo criterio de optimalidad, Φ , es a través de la razón

$$\Phi - \text{ef}(\xi) = \frac{\Phi(\xi)}{\Phi(\xi^*)}. \quad (2-33)$$

Esta medida toma valores entre 0 y 1, y cuanto más cerca esté de 1 mejor será la eficiencia. Para el criterio de D-optimalidad, la D-eficiencia se define como:

$$D - \text{ef}(\xi) = \left[\frac{\det M(\xi)}{\det M(\xi_D^*)} \right]^{1/p}. \quad (2-34)$$

Este valor permite determinar la capacidad del diseño ξ para estimar de manera eficiente los parámetros del modelo en comparación con el diseño ξ_D^* .

En el caso de T-optimalidad:

$$T - \text{ef}(\xi) = \frac{\Delta(\xi)}{\Delta(\xi_T^*)}, \quad (2-35)$$

donde ξ_T^* es el diseño T-óptimo y $\Delta(\xi)$ está dado por (2-25).

3. Diseños óptimos para discriminación entre modelos marginales mixtos tipo I

En este capítulo se propone una metodología para el problema de discriminación en una clase de modelos no lineales de efectos mixtos. El criterio de diseño propuesto se desarrolla con argumentos teóricos similares a los utilizados en la interpretación estadística del criterio de T-optimalidad, introducido por Atkinson y Fedorov (1975), pero en el contexto de los modelos mixtos. El resultado más importante del capítulo es la formulación del teorema de equivalencia para el criterio de discriminación mediante la aplicación de la teoría de diseños aproximados.

En general, un modelo no lineal de efectos mixtos se puede definir como un modelo donde la función de respuesta depende de algunos efectos fijos y otros aleatorios. Una clase especial de estos modelos son aquellos donde la función de respuesta es lineal en los efectos aleatorios, llamados modelos no lineales marginales (ver Demidenko, 2004). En un modelo marginal tanto la media como la matriz de varianzas y covarianzas tienen una expresión cerrada, haciendo el proceso de estimación de los parámetros del modelo menos complejo. El modelo no lineal con coeficientes aleatorios y el modelo linealizado de un modelo de efectos mixtos son ejemplos de modelos marginales. De acuerdo con la estructura de la matriz de diseño de los efectos aleatorios, se pueden definir dos clases de modelos marginales: (a) los modelos marginales tipo I en los que la matriz de diseño es constante y no depende del vector de parámetros poblacionales, y (b) los modelos marginales tipo II donde la matriz de diseño depende del vector de parámetros poblacionales. Estos modelos aunque en estructura son muy similares tienen propiedades estadísticas diferentes (ver Vonesh y Chinchilli, 1997; Demidenko, 2004).

En este capítulo el objetivo es construir diseños para discriminación entre dos modelos marginales tipo I anidados, los cuales difieren en su matriz de covarianzas intra-individuos.

El capítulo se desarrolla en cuatro secciones: en la sección 3.1 se describe el modelo considerado, en la sección 3.2 se presenta la definición de diseño poblacional, en la sección 3.3 se describe el problema de discriminación entre modelos, en la sección 3.4 se muestra una generalización del criterio de T-optimalidad para modelos marginales tipo I, en la sección 3.5 se enuncia y se demuestra el respectivo teorema de equivalencia, y finalmente, en la sección 3.6 se ilustra la metodología propuesta con dos ejemplos de aplicación.

3.1. Modelo

Suponga que se tienen N sujetos y cada sujeto se observa n veces a través del tiempo sin repeticiones en cada tiempo de medición. Sea $\mathbf{y}_i = (y_{i1}, \dots, y_{in})^T$ el vector de observaciones del i -ésimo individuo y sea $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^T$ el vector que contiene los diferentes tiempos en los que se realizan dichas observaciones, con x_{ij} perteneciente a un conjunto finito $\mathcal{X} \subset \mathbb{R}$ conocido. Suponga además que las observaciones en individuos distintos son independientes. Para describir la relación entre los vectores \mathbf{y}_i y \mathbf{x}_i se considera el modelo no lineal de efectos mixtos:

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N \quad (3-1)$$

donde:

- $\boldsymbol{\beta}$ es un vector p -dimensional de parámetros poblacionales fijos desconocidos.
- $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta})$ es un vector n -dimensional tal que $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) = (f(x_{i1}, \boldsymbol{\beta}), \dots, f(x_{in}, \boldsymbol{\beta}))^T$ donde $f(x, \boldsymbol{\beta})$ es una función conocida no lineal del vector de parámetros $\boldsymbol{\beta}$.
- \mathbf{Z}_i es una matriz de orden $n \times q$ de rango completo y es la matriz de diseño asociada a los efectos aleatorios.
- \mathbf{b}_i es un vector q -dimensional de efectos aleatorios asociados con el i -ésimo individuo. Se supone que los vectores $\mathbf{b}_1, \dots, \mathbf{b}_N$ son independientes con distribución $\mathcal{N}_q(\mathbf{0}, \mathbf{D})$ donde \mathbf{D} es una matriz definida positiva común para todos los individuos.
- $\boldsymbol{\varepsilon}_i$ es el vector de errores aleatorios. Se supone que los vectores $\boldsymbol{\varepsilon}_i$ son independientes entre sí y de los vectores \mathbf{b}_i con distribución $\mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}))$ donde $\sigma^2 > 0$ y $\boldsymbol{\lambda} \in \Omega \subset \mathbb{R}^d$ es un vector que no depende de $\boldsymbol{\beta}$. La estructura de la matriz $\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})$ es la misma para todos los individuos y varía de un sujeto a otro solamente a través de los valores diferentes del vector \mathbf{x}_i .

El modelo (3-1) tiene, entre otras, las siguientes características (ver Vonesh y Chinchilli, 1997; Demidenko, 2004):

- (i) El modelo lineal de efectos mixtos y el modelo de curvas de crecimiento con coeficientes aleatorios son casos especiales del modelo (3-1).
- (ii) La matriz \mathbf{Z}_i es fija y no depende del vector de parámetros $\boldsymbol{\beta}$.
- (iii) El modelo es lineal en los efectos aleatorios.
- (iv) Si \mathbf{Z}_i es igual al vector de unos $\mathbf{1}_q = (1, \dots, 1)^T$ se obtiene la estructura de simetría compuesta en modelos univariados de medidas repetidas con efectos mixtos univariados.

- (v) El modelo puede ser modificado para incluir el modelo no lineal de efectos mixtos de Beal y Sheiner (1982).
- (vi) Este modelo es llamado modelo marginal debido a que la media marginal de \mathbf{y}_i se puede expresar en forma cerrada como una función del vector de parámetros poblacionales, esto es, $E(\mathbf{y}_i) = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta})$. Siguiendo Demidenko (2004) este modelo es llamado *modelo marginal tipo I*.

Sea $\boldsymbol{\psi}$ el vector que contiene los $q(q+1)/2$ elementos diferentes de la matriz \mathbf{D} y sea $\boldsymbol{\gamma} = (\sigma^2, \boldsymbol{\psi}^T, \boldsymbol{\lambda}^T)^T \in \Gamma \subset \mathbb{R}^m$ el vector que contiene las componentes de varianza. Bajo las formulaciones anteriores, se tiene que la distribución marginal de \mathbf{y}_i es normal n -variada con vector de medias y matriz de varianzas y covarianzas dadas por

$$\begin{aligned} E(\mathbf{y}_i) &= \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) \\ Cov(\mathbf{y}_i) &= \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) \\ &\equiv \boldsymbol{\Sigma}(\mathbf{x}_i, \boldsymbol{\gamma}). \end{aligned} \quad (3-2)$$

Este modelo también se puede escribir como

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}(\mathbf{x}_i, \boldsymbol{\gamma})). \quad (3-3)$$

A partir de (3-2) se observa que el vector de medias y la matriz de covarianzas no tienen parámetros comunes, es decir, $\boldsymbol{\beta}$ y $\boldsymbol{\gamma}$ no se superponen. Para la estimación del vector de parámetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ se pueden aplicar diferentes enfoques basados en el método de máxima verosimilitud. Información más detallada acerca del problema de estimación en modelos marginales se encuentra, por ejemplo, en Vonesh y Chinchilli (1997) y Demidenko (2004), entre otros.

Estructuras de covarianza

La matriz \mathbf{D} representa la variabilidad entre los individuos y es llamada matriz de covarianzas entre-individuos. En algunas aplicaciones es común, en la práctica, no asumir una estructura particular para la matriz \mathbf{D} excepto en aquellos casos donde consideraciones previas sugieren un modelo de covarianzas particular (ver Davidian y Giltinan, 1995; Hand y Crowder, 1996). La variabilidad dentro de un individuo está representada por la matriz $\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})$, conocida como la matriz de covarianzas intra-individuos. Debido a que las observaciones en un mismo individuo se miden a través del tiempo, la matriz \mathbf{R} puede tener distintas parametrizaciones que dan cuenta de la heteroscedasticidad de los errores y la estructura de correlación entre ellos. Algunos de los modelos más utilizados para la estructura de covarianzas intra-individuo son:

- (i) Estructura de independencia, asume errores no correlacionados con varianzas iguales

$$\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = \sigma^2 \mathbf{I}_n$$

donde \mathbf{I}_n es la matriz identidad de orden $n \times n$.

- (ii) Varianza constante y un modelo autorregresivo de primer orden para la estructura de correlación

$$\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = \sigma^2 \mathbf{W}(\mathbf{x}_i, \rho) = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}, \quad (3-4)$$

donde $\boldsymbol{\lambda} = \rho$ y $-1 < \rho < 1$.

- (iii) Estructura de simetría compuesta

$$\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = \sigma^2 [(1 - \rho)\mathbf{I}_n + \rho\mathbf{J}_n]$$

donde \mathbf{J}_n es una matriz de unos de orden n y $\boldsymbol{\lambda} = \rho$.

- (iv) Errores no correlacionados y heteroscedásticos con función de varianza $g(x, \boldsymbol{\alpha})$ donde

$$\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = G(\mathbf{x}_i, \sigma^2, \boldsymbol{\alpha}) = \sigma^2 \text{diag}(g(x_{i1}, \boldsymbol{\alpha}), \dots, g(x_{in}, \boldsymbol{\alpha})) \quad (3-5)$$

aquí $\boldsymbol{\lambda} = \boldsymbol{\alpha}$.

- (v) Modelos que combinan una función de varianza con una estructura de correlación. Por ejemplo, modelo de varianza exponencial, útil en aquellos casos donde la varianza tiende a aumentar o a disminuir a través del tiempo en forma exponencial, y el supuesto de un modelo autorregresivo para la estructura de correlación:

$$\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = \begin{bmatrix} g_{i1}^2 & \rho g_{i1} g_{i2} & \dots & \rho^{n-1} g_{i1} g_{in} \\ \rho g_{i1} g_{i2} & g_{i2}^2 & \dots & \rho^{n-2} g_{i2} g_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} g_{i1} g_{in} & \rho^{n-2} g_{i2} g_{in} & \dots & g_{in}^2 \end{bmatrix}, \quad (3-6)$$

donde $g_{ij} = \exp(\alpha x_{ij})$ y $\boldsymbol{\lambda} = (\alpha, \rho)^T$.

Otras estructuras de covarianza se pueden consultar en Pinheiro y Bates (2000).

En general, para un individuo dado, la matriz \mathbf{R} se puede expresar como:

$$\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = \mathbf{G}^{1/2}(\mathbf{x}_i, \sigma^2, \boldsymbol{\alpha}) \mathbf{W}(\mathbf{x}_i, \rho) \mathbf{G}^{1/2}(\mathbf{x}_i, \sigma^2, \boldsymbol{\alpha}), \quad (3-7)$$

donde $\mathbf{G}(\mathbf{x}_i, \sigma^2, \boldsymbol{\alpha})$ es una matriz diagonal de orden $(n \times n)$ que caracteriza el modelo para las varianzas y $\mathbf{W}(\mathbf{x}_i, \rho)$ es una matriz de orden $(n \times n)$ que describe el modelo de correlación, con $\boldsymbol{\lambda} = (\boldsymbol{\alpha}^T, \rho^T)^T$.

3.2. Diseño poblacional

En el caso de modelos de efectos mixtos un diseño experimental se puede definir en dos etapas: en la primera etapa se especifica el diseño a nivel individual y en la segunda se define el diseño para el grupo total de individuos (ver Patan y Bogacka, 2010).

Definición 3.2.1. (Nivel individual) Para cada individuo i , $i = 1, \dots, N$, un diseño individual ξ_i con n_i observaciones es un conjunto de condiciones experimentales x_{i1}, \dots, x_{ik_i} junto con el número de repeticiones n_{i1}, \dots, n_{ik_i} en cada punto x_{ij} tal que $\sum_{j=1}^{k_i} n_{ij} = n_i$. En forma resumida el diseño individual ξ_i se puede representar como

$$\xi_i = \left\{ \begin{array}{ccc} x_{i1} & \cdots & x_{ik_i} \\ w_{i1} & \cdots & w_{ik_i} \end{array} \right\} = \{\mathbf{x}_{ij}, w_{ij}\}_{j=1}^{k_i}, \quad w_{ij} = n_{ij}/n_i, \quad \sum_{j=1}^{k_i} n_{ij} = n_i. \quad (3-8)$$

De acuerdo con (2-8) el diseño individual ξ_i es un diseño exacto normalizado y el correspondiente diseño aproximado está dado por

$$\xi_i = \{\mathbf{x}_{ij}, w_{ij}\}_{j=1}^{k_i}, \quad w_{ij} \in [0, 1], \quad \sum_{j=1}^{k_i} w_{ij} = 1. \quad (3-9)$$

Definición 3.2.2. (Nivel de grupo) Para una muestra dada de N individuos, un diseño poblacional ζ es un conjunto de diseños individuales ξ_1, \dots, ξ_s junto con el número de individuos que serán observados bajo cada diseño individual N_1, \dots, N_s tal que $\sum_{k=1}^s N_k = N$. En forma resumida un diseño poblacional ζ se puede representar como

$$\zeta = \left\{ \begin{array}{ccc} \xi_1 & \cdots & \xi_s \\ N_1 & \cdots & N_s \end{array} \right\} = \{\xi_k, N_k\}_{k=1}^s, \quad \sum_{k=1}^s N_k = N. \quad (3-10)$$

Observaciones 3.2.1. Las siguientes son algunas observaciones con respecto a la definición de diseño poblacional:

1. Si los diseños individuales ξ_i son exactos entonces el diseño poblacional

$$\{\xi_k, \omega_k\}_{k=1}^s, \quad \omega_k = N_k/N, \quad \sum_{k=1}^s N_k = N \quad (3-11)$$

es llamado *diseño poblacional exacto*.

2. El conjunto de individuos con el mismo diseño individual ξ_k conforman un grupo.
3. Si todos los individuos son observados bajo el mismo diseño individual entonces se dice que el diseño poblacional es de un solo grupo.

Bajo las formulaciones del modelo (3-1), los vectores n -dimensionales \mathbf{x}_i de tiempos en los que las observaciones se realizarán para cada individuo son diseños individuales ξ_i exactos sin repeticiones en cada punto de observación. Así, para una muestra de N individuos, bajo el modelo (3-1) un diseño poblacional exacto está dado por

$$\zeta_N = \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_s \\ \omega_1 & \cdots & \omega_s \end{array} \right\} = \{\mathbf{x}_k, \omega_k\}_1^s, \quad \mathbf{x}_k \in \bar{\mathcal{X}}^n, \quad \omega_k = \frac{N_k}{N} \quad (3-12)$$

donde $\bar{\mathcal{X}}^n = \{(x_1, \dots, x_n) : x_j \in \mathcal{X}\}$ es el conjunto que contiene todas las combinaciones sin repetición de elementos de \mathcal{X} tomados de a n , el cual se supone conocido. Se denotará por r el número de elementos de $\bar{\mathcal{X}}^n$.

El problema de diseño

Para valores fijos del número de individuos N y el número de tiempos de medición n , el problema de diseño poblacional óptimo consiste en determinar los diferentes vectores $\mathbf{x}_i \in \bar{\mathcal{X}}^n$ y el número de sujetos observados bajo cada vector \mathbf{x}_i de tal forma que el diseño obtenido maximice algún criterio de optimalidad cuya elección dependerá del objetivo del experimento, es decir, elegir el diseño que verifica

$$\zeta_N^* = \arg \max_{\zeta_N} \Phi(\zeta_N), \quad (3-13)$$

donde Φ es una función acorde con el propósito del experimento como se describió en la sección 2.3. Debido a que varios de estos criterios generalmente dependen de los parámetros del modelo, un enfoque comúnmente adoptado es hallar diseños óptimos locales. Este enfoque consiste en especificar un valor a priori del vector de parámetros y realizar el proceso de optimización para ese valor específico (Chernoff, 1953).

Diseño poblacional aproximado

El problema de optimización (3-13) es un problema de optimización entera

$$\zeta_N^* = \{\mathbf{x}_k^*, \omega_k^*\}_1^{s^*} = \arg \max_{\mathbf{x}_k, \omega_k, s} \Phi(\{\mathbf{x}_k, \omega_k\}_1^s), \quad N\omega_k \text{ entero} \quad (3-14)$$

el cual, en general es más difícil de resolver. Con el fin de simplificar este problema y proporcionar condiciones necesarias para la optimalidad de un diseño con respecto a algún criterio de optimalidad, se define el correspondiente diseño poblacional aproximado.

Asumiendo que los N_k son valores moderadamente grandes, un diseño poblacional aproximado está dado por

$$\zeta = \{\mathbf{x}_k, \omega_k\}_1^s, \quad \mathbf{x}_k \in \bar{\mathcal{X}}^n, \quad 0 \leq \omega_k \leq 1, \quad \sum_{k=1}^s \omega_k = 1 \quad (3-15)$$

donde los pesos ω_i representan la proporción del total de individuos a ser observados en el punto \mathbf{x}_i .

El soporte del diseño ζ está dado por

$$\text{Supp}(\zeta) = \{\mathbf{x}_k \in \bar{\mathcal{X}}^n | \omega_k > 0\} = \{\mathbf{x}_1, \dots, \mathbf{x}_s\}. \quad (3-16)$$

A partir de la definición de los pesos ω_k en (3-15), se tiene que cualquier medida de probabilidad definida en el espacio de diseño finito $\bar{\mathcal{X}}^n$ es un diseño. Así, un diseño aproximado se puede especificar por medio del vector de pesos $\boldsymbol{\omega} \in \Xi$ donde

$$\Xi = \left\{ (\omega_1, \dots, \omega_r) : \omega_i \geq 0, i = 1, \dots, r \text{ y } \sum_{i=1}^r \omega_i = 1 \right\}. \quad (3-17)$$

Valores iguales a cero en algunas componentes del vector $\boldsymbol{\omega}$ indicarían que el vector de tiempos correspondiente no es considerado en el experimento.

Así el problema de optimización (3-13), en el contexto de los diseños aproximados, consiste en determinar el diseño óptimo que verifica

$$\boldsymbol{\omega}^* = \arg \max_{\boldsymbol{\omega} \in \Xi} \Phi(\boldsymbol{\omega}). \quad (3-18)$$

Nótese, que si $\boldsymbol{\omega}_1$ y $\boldsymbol{\omega}_2$ son medidas de probabilidad definidas en $\bar{\mathcal{X}}^n$ y $\alpha \in [0, 1]$ entonces $\boldsymbol{\omega} = (1 - \alpha)\boldsymbol{\omega}_1 + \alpha\boldsymbol{\omega}_2$ también es una medida de probabilidad en $\bar{\mathcal{X}}^n$, por tanto el conjunto de todos los diseños aproximados Ξ es un conjunto convexo.

En lo que sigue se hace referencia a los diseños aproximados $\boldsymbol{\omega}$ localmente óptimos.

3.3. Discriminación entre modelos marginales mixtos

En esta sección se describe el problema de discriminación considerado en este trabajo.

Asuma que la función de respuesta f y la matriz de covarianzas entre-individuos \mathbf{D} bajo el modelo (3-1) se especifican de forma correcta. Como se mencionó en la sección 3.1, la matriz que describe la variabilidad de las observaciones en un mismo individuo puede tener diferentes estructuras dependiendo del problema de estudio. Bajo los supuestos del modelo (3-1), suponga que $\mathbf{R}_1(\mathbf{x}_i, \sigma_1^2, \boldsymbol{\lambda}_1)$ y $\mathbf{R}_2(\mathbf{x}_i, \sigma_2^2, \boldsymbol{\lambda}_2)$ son dos modelos alternativos para la matriz de covarianzas intra-individuos con la misma estructura dada por

$$\mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = \mathbf{G}^{1/2}(\mathbf{x}_i, \sigma^2, \boldsymbol{\alpha})\mathbf{W}(\mathbf{x}_i, \boldsymbol{\rho})\mathbf{G}^{1/2}(\mathbf{x}_i, \sigma^2, \boldsymbol{\alpha}), \quad \boldsymbol{\lambda} = (\boldsymbol{\alpha}^T, \boldsymbol{\rho}^T)^T. \quad (3-19)$$

Suponga, además, que el modelo \mathbf{R}_2 está anidado dentro de \mathbf{R}_1 con respecto al vector de parámetros $\boldsymbol{\lambda}$. Esto significa que el espacio paramétrico Ω_2 de \mathbf{R}_2 es un subconjunto del espacio de parámetros Ω_1 de \mathbf{R}_1 definido por κ restricciones de igualdad a partir de funciones continuamente diferenciables, es decir,

$$\Omega_2 = \{\boldsymbol{\lambda} \in \Omega_1 : h_\tau(\boldsymbol{\lambda}) = 0, \tau = 1, \dots, \kappa\}$$

donde las funciones $h_\tau(\boldsymbol{\lambda})$ son continuamente diferenciables. El objetivo es construir diseños poblacionales óptimos que permitan decidir cuál de los dos modelos es el más adecuado.

En la revisión bibliográfica realizada, este problema de diseño no ha sido considerado y aunque algunas metodologías, como las que se describen a continuación, han sido sugeridas con el fin de discriminar entre dos modelos de efectos mixtos, éstas pueden ser poco eficientes o no se pueden aplicar directamente en el caso considerado.

Waterhouse *et al.* (2005) sugirieron el criterio producto de D-optimalidad basado en el producto de los determinantes de las matrices de información de Fisher para estimación y discriminación simultáneamente. Sin embargo, para modelos no lineales, estos diseños pueden ser menos eficientes para discriminar en comparación con los diseños T-óptimos.

Vajjah y Duffull (2012) propusieron una metodología para encontrar diseños T-óptimos robustos que no depende de la selección a priori del modelo correcto. Sin embargo, en el caso de dos modelos de variación anidados esta metodología no puede ser aplicada debido a que los diseños T-óptimos que se derivan allí se basan sólo en los modelos de efectos fijos sin error residual.

Kuczewski *et al.* (2008) propusieron una extensión del criterio de T-optimalidad para modelos heteroscedásticos con múltiples respuestas pero este enfoque sólo se aplica directamente para el caso de diseños de un solo grupo.

En este trabajo se propone una generalización del criterio de T-optimalidad con el fin de discriminar entre dos modelos de covarianzas intra-individuos anidados $\mathbf{R}_1(\mathbf{x}_i, \sigma_1^2, \boldsymbol{\lambda}_1)$ y $\mathbf{R}_2(\mathbf{x}_i, \sigma_2^2, \boldsymbol{\lambda}_2)$. Se muestra que los diseños que maximizan este criterio proporcionan el test de la razón de verosimilitud más potente cuando el modelo más grande se asume como el modelo correcto y el correspondiente vector de parámetros es completamente conocido. En la siguiente sección se define el criterio de optimalidad para discriminación.

3.4. Criterio propuesto para dicriminación entre dos modelos intra-individuales

El problema de discriminación entre dos modelos de covarianzas intra-individuos anidados descrito previamente conduce al problema de discriminar entre dos modelos marginales \mathcal{M}_1 y \mathcal{M}_2 con vectores de parámetros $\boldsymbol{\theta}_1 \in \Theta_1 \subset \mathbb{R}^{v_1}$ y $\boldsymbol{\theta}_2 \in \Theta_2 \subset \mathbb{R}^{v_2}$, los cuales asumen la misma estructura tanto para la función de respuesta media $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta})$ como para la matriz de covarianzas entre-individuos \mathbf{D} , pero suponen modelos diferentes para la variabilidad intra-individual. En forma compacta estos modelos se pueden representar como:

$$\begin{aligned} \mathcal{M}_1 : \varepsilon_i &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}_1(\mathbf{x}_i, \sigma_1^2, \boldsymbol{\lambda}_1)), \quad \boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1^T, \boldsymbol{\gamma}_1^T)^T \in \Theta_1 \\ \boldsymbol{\gamma}_1 &= (\sigma_1^2, \boldsymbol{\psi}_1^T, \boldsymbol{\lambda}_1^T)^T \in \Gamma_1 \subset \mathbb{R}^{m_1} \end{aligned}$$

$$\begin{aligned} \mathcal{M}_2 : \varepsilon_i &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}_2(\mathbf{x}_i, \sigma_2^2, \boldsymbol{\lambda}_2)), \quad \boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2^T, \boldsymbol{\gamma}_2^T)^T \in \Theta_2 \\ \boldsymbol{\gamma}_2 &\in \Gamma_2 = \{\boldsymbol{\gamma}_2 = (\sigma_2^2, \boldsymbol{\psi}_2^T, \boldsymbol{\lambda}_2^T)^T \in \Gamma_1 : h_\tau(\boldsymbol{\lambda}_2) = 0, \tau = 1, \dots, \kappa\}. \end{aligned}$$

Con el fin de discriminar entre \mathcal{M}_1 y \mathcal{M}_2 se propone el siguiente criterio de optimalidad.

Definición 3.4.1. *Asumiendo que \mathcal{M}_1 es el modelo correcto con vector de parámetros $\boldsymbol{\gamma}_1^0$ conocido, un diseño poblacional óptimo local ζ_N^* para discriminar entre \mathcal{M}_1 y \mathcal{M}_2 verifica*

$$\zeta_N^* = \arg \max_{\zeta_N} T_W^I(\zeta_N) \quad (3-20)$$

donde

$$T_W^I(\zeta_N) = \min_{\boldsymbol{\gamma}_2 \in \Gamma_2} \sum_{k=1}^s \frac{N_k}{N} F(\mathbf{x}_k, \boldsymbol{\gamma}_2) \quad (3-21)$$

con

$$F(\mathbf{x}, \boldsymbol{\gamma}_2) = \text{tr}[\boldsymbol{\Sigma}(\mathbf{x})\boldsymbol{\Sigma}_2^{-1}(\mathbf{x}, \boldsymbol{\gamma}_2)] - \log \det[\boldsymbol{\Sigma}(\mathbf{x})\boldsymbol{\Sigma}_2^{-1}(\mathbf{x}, \boldsymbol{\gamma}_2)] \quad (3-22)$$

y $\boldsymbol{\Sigma}(\mathbf{x}) = \boldsymbol{\Sigma}_1(\mathbf{x}, \boldsymbol{\gamma}_1^0)$.

El diseño ζ_N^* es llamado T_W^I -óptimo, donde la letra W se refiere a la variabilidad dentro (“Within”) y el superíndice se refiere al tipo de modelo marginal. Este diseño es óptimo local pues el criterio depende del valor del vector de parámetros $\boldsymbol{\gamma}_1$.

A continuación se presenta la justificación estadística del criterio T_W^I -optimal.

3.4.1. Justificación del criterio

Para discriminar entre los modelos alternativos \mathcal{M}_1 y \mathcal{M}_2 considere el test de la razón de verosimilitud usual para comparación de modelos. Sea $\boldsymbol{\gamma}^0$ el valor verdadero pero desconocido del vector de parámetros $\boldsymbol{\gamma}$ y considere el problema de contraste:

$$H_0 : \boldsymbol{\gamma}^0 \in \Gamma_2 \quad \text{versus} \quad H_1 : \boldsymbol{\gamma}^0 \in \Gamma_1, \quad (3-23)$$

donde $\Gamma_2 = \{\boldsymbol{\gamma} = (\sigma^2, \boldsymbol{\psi}^T, \boldsymbol{\lambda}^T)^T | \boldsymbol{\gamma} \in \Gamma_1, h_\tau(\boldsymbol{\lambda}) = 0, \tau = 1, \dots, \kappa\}$.

Los siguientes supuestos son establecidos:

(A1) Γ_1 es un conjunto compacto,

(A2) $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta})$ es una función continua y dos veces continuamente diferenciable en Γ_1

(A3) $\boldsymbol{\Sigma}(\mathbf{x}, \boldsymbol{\gamma})$ es una función continua y dos veces continuamente diferenciable en Γ_1 .

Sean $\mathbf{y}_{k1}, \dots, \mathbf{y}_{kN_k}$ vectores de observaciones independientes con vector de tiempos de medición \mathbf{x}_k , $k = 1, \dots, s$. Bajo el modelo marginal (3-1), $\mathbf{y}_{k1}, \dots, \mathbf{y}_{kN_k}$, constituye una muestra aleatoria de vectores que provienen de una distribución $N_n(\mathbf{f}(\mathbf{x}_k, \boldsymbol{\beta}), \boldsymbol{\Sigma}(\mathbf{x}_k, \boldsymbol{\gamma}))$ donde $\boldsymbol{\mu}_k = \mathbf{f}(\mathbf{x}_k, \boldsymbol{\beta})$. Debido a que el interés está sólo en las estructuras de covarianzas $\boldsymbol{\Sigma}(\mathbf{x}_k, \boldsymbol{\gamma})$ y los vectores $\boldsymbol{\beta}$ y $\boldsymbol{\gamma}$ no se superponen, los vectores $\boldsymbol{\mu}_k$ son considerados parámetros molestos. Para el problema de contraste (3-23), la función de verosimilitud con base en s muestras independientes está dada por

$$L(\boldsymbol{\gamma}, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_s) = \prod_{k=1}^s (2\pi)^{-N_k n/2} (\det \boldsymbol{\Sigma}(\mathbf{x}_k, \boldsymbol{\gamma}))^{-N_k/2} \text{etr} \left[-\frac{1}{2} \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k, \boldsymbol{\gamma}) \mathbf{A}_k \right] \\ \times \exp \left[-\frac{1}{2} N_k (\bar{\mathbf{y}}_k - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k, \boldsymbol{\gamma}) (\bar{\mathbf{y}}_k - \boldsymbol{\mu}_k) \right],$$

donde $\bar{\mathbf{y}}_k = N_k^{-1} \sum_{m=1}^{N_k} \mathbf{y}_{km}$, $\mathbf{A}_k = \sum_{m=1}^{N_k} (\mathbf{y}_{km} - \bar{\mathbf{y}}_k)(\mathbf{y}_{km} - \bar{\mathbf{y}}_k)^T$ y $\text{etr}(\cdot) = \exp[\text{tr}(\cdot)]$. El logaritmo de esta función es

$$\ell(\boldsymbol{\gamma}, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_s) = \sum_{k=1}^s N_k \left[-\frac{n}{2} \log 2\pi - \frac{1}{2} \log \det(\boldsymbol{\Sigma}(\mathbf{x}_k, \boldsymbol{\gamma})) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{x}_k, \boldsymbol{\gamma}) \mathbf{S}_k) \right. \\ \left. - \frac{1}{2} (\bar{\mathbf{y}}_k - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k, \boldsymbol{\gamma}) (\bar{\mathbf{y}}_k - \boldsymbol{\mu}_k) \right],$$

donde $\mathbf{S}_k = N_k^{-1} \mathbf{A}_k$.

Si $\hat{\boldsymbol{\gamma}}_1$ y $\hat{\boldsymbol{\gamma}}_2$ son los estimadores de máxima verosimilitud sobre Γ_1 y Γ_2 respectivamente, y $\hat{\boldsymbol{\mu}}_k = \bar{\mathbf{y}}_k$, entonces el estadístico de la razón de verosimilitud para (3-23), está dado por

$$-2 \log \hat{\lambda}_N = \sum_{k=1}^s N_k \{ F_{ML}(\mathbf{S}_k, \boldsymbol{\Sigma}_2(\mathbf{x}_k, \hat{\boldsymbol{\gamma}}_2)) - F_{ML}(\mathbf{S}_k, \boldsymbol{\Sigma}_1(\mathbf{x}_k, \hat{\boldsymbol{\gamma}}_1)) \}, \quad (3-24)$$

donde

$$F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\mathbf{x}, \boldsymbol{\gamma})) = \log \det(\boldsymbol{\Sigma}(\mathbf{x}, \boldsymbol{\gamma})) + \text{tr} [\boldsymbol{\Sigma}^{-1}(\mathbf{x}, \boldsymbol{\gamma}) \mathbf{S}] - \log \det(\mathbf{S}) - n. \quad (3-25)$$

Bajo H_0 , el estadístico de prueba $-2 \log \hat{\lambda}_N$ sigue asintóticamente una distribución χ^2 con κ grados de libertad. Por tanto, un test aproximado basado en $-2 \log \hat{\lambda}_N$ de tamaño α rechaza H_0 si $-2 \log \hat{\lambda}_N > c_\kappa(\alpha)$, donde $c_\kappa(\alpha)$ denota el punto porcentual $100\alpha\%$ superior de la distribución χ_κ^2 .

En el análisis de modelos de estructuras de covarianzas, la función $F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\mathbf{x}, \boldsymbol{\gamma}))$ es conocida como la función de discrepancia de máxima verosimilitud (ver Steiger *et al.*, 1985).

La extensión de esta función en el caso de s grupos está definida por (ver Bollen, 1989):

$$F(\boldsymbol{\gamma}) = \sum_{k=1}^s \frac{N_k}{N} F_{ML}(\mathbf{S}_k, \boldsymbol{\Sigma}(\mathbf{x}_k, \hat{\boldsymbol{\gamma}})). \quad (3-26)$$

Así, el estadístico (3-24) se puede escribir como

$$-2 \log \widehat{\lambda}_N = N(F(\widehat{\gamma}_2) - F(\widehat{\gamma}_1)). \quad (3-27)$$

Ahora, la función potencia del test está dada por $P(-2 \log \widehat{\lambda}_N > c_\kappa(\alpha) \mid \gamma \in \Gamma_1)$, la cual depende del valor del vector de parámetros bajo la hipótesis alternativa. Para un valor dado de N denotado por N_0 , y un valor específico $\gamma_1^0 \in \Gamma_1$ cercano a Γ_2 , esta probabilidad se puede aproximar considerando la distribución asintótica de $-2 \log \widehat{\lambda}_N$ bajo una sucesión $\{\gamma_N^0\}$ de alternativas locales que convergen a un punto interior γ_2^0 en Γ_1 bajo H_0 (ver Stroud, 1972). Debido a que $F(\gamma)$ en (3-26) es una función de discrepancia, bajo los supuestos (A1)-(A3) y condiciones de regularidad, esta distribución asintótica es una distribución chi-cuadrado no centrada $\chi_\kappa^2(\delta)$ con κ grados de libertad y parámetro de no centralidad δ , el cual se puede aproximar mediante el valor:

$$\begin{aligned} \tilde{\delta} &= N \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^s \frac{N_k}{N} F_{ML}(\Sigma_1(\mathbf{x}_k, \gamma_1^0), \Sigma_2(\mathbf{x}_k, \gamma_2)) \\ &= N \times \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^s \frac{N_k}{N} F(\mathbf{x}_k, \gamma_2) \\ &= N \times T_W^I(\zeta_N), \end{aligned} \quad (3-28)$$

(ver Satorra y Saris, 1985; Steiger *et al.*, 1985). Debido a que la potencia de la prueba es una función monótona creciente del parámetro de no centralidad, a partir de (3-28) se tiene que la potencia es una función creciente de $T_W^I(\zeta_N)$ y por tanto se maximiza mediante la elección del diseño ζ_N .

3.4.2. Diseños T_W^I -óptimos aproximados

Debido a que el objetivo principal es caracterizar los diseños T_W^I -óptimos mediante la aplicación de la teoría de diseños aproximados, se define el diseño T_W^I -óptimo en el contexto de diseños poblaciones aproximados.

Definición 3.4.2. *Asumiendo que \mathcal{M}_1 es el modelo correcto con vector de parámetros γ_1^0 conocido, un diseño poblacional aproximado T_W^I -óptimo local ω^* para discriminar entre \mathcal{M}_1 y \mathcal{M}_2 verifica*

$$\omega^* = \arg \max_{\omega \in \Xi} T_W^I(\omega) \quad (3-29)$$

donde

$$T_W^I(\omega) = \min_{\gamma_2 \in \Gamma_2} \sum_{i=1}^r \omega_i F(\mathbf{x}_i, \gamma_2) \quad (3-30)$$

con $F(\mathbf{x}_i, \gamma_2)$ dada por (3-22).

En la siguiente sección se presenta el principal resultado de este capítulo.

3.5. Teorema de Equivalencia

De acuerdo con la aplicación de la teoría de diseños aproximados dos condiciones son necesarias en la formulación del teorema de equivalencia para un criterio de optimalidad: la concavidad del criterio y una expresión para la derivada direccional. Los siguientes resultados establecen estas dos propiedades para el criterio de T_W^I -optimalidad. Así como en el caso de T-optimalidad, con el fin de simplificar el chequeo de un diseño T_W^I -óptimo, se adopta la siguiente definición de diseño regular.

Definición 3.5.1. *Un diseño ω es llamado T_W^I -regular si el conjunto*

$$\Gamma_2(\omega) = \left\{ \tilde{\gamma}_2 : \tilde{\gamma}_2(\omega) = \arg \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^r \omega_k F(\mathbf{x}_k, \gamma_2) \right\}$$

es unitario, de lo contrario es llamado T_W^I -singular.

Nótese que si ω es T_W^I -regular y $\tilde{\gamma}_2 \in \Gamma_2(\omega)$ entonces $\tilde{\gamma}_2$ es la única solución de la ecuación

$$\sum_{k=1}^r \omega_k F(\mathbf{x}_k, \tilde{\gamma}_2) = \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^r \omega_k F(\mathbf{x}_k, \gamma_2).$$

En los siguientes teoremas se demuestra la concavidad de la función T_W^I en el conjunto Ξ y la existencia de la derivada direccional.

Teorema 3.5.1. *Sean $\omega_1, \omega_2 \in \Xi$ y $\alpha \in [0, 1]$. Entonces*

$$T_W^I((1 - \alpha)\omega_1 + \alpha\omega_2) \geq (1 - \alpha)T_W^I(\omega_1) + \alpha T_W^I(\omega_2). \quad (3-31)$$

Demostración. Sea $\omega = (1 - \alpha)\omega_1 + \alpha\omega_2$, entonces

$$\begin{aligned} T_W^I(\omega) &= \min_{\gamma_2 \in \Gamma_2} \left[(1 - \alpha) \sum_{i=1}^r \omega_{1i} F(\mathbf{x}_i, \gamma_2) + \alpha \sum_{i=1}^r \omega_{2i} F(\mathbf{x}_i, \gamma_2) \right] \\ &\geq (1 - \alpha) \min_{\gamma_2 \in \Gamma_2} \sum_{i=1}^r \omega_{1i} F(\mathbf{x}_i, \gamma_2) + \alpha \min_{\gamma_2 \in \Gamma_2} \sum_{i=1}^r \omega_{2i} F(\mathbf{x}_i, \gamma_2) \\ &= (1 - \alpha)T_W^I(\omega_1) + \alpha T_W^I(\omega_2). \end{aligned}$$

Por tanto $T_W^I(\omega)$ es una función cóncava en Ξ . □

Teorema 3.5.2. *Sean ω y $\bar{\omega}$ diseños poblacionales con ω un diseño T_W^I -regular. Entonces, bajo los supuestos (A1)-(A3), la derivada direccional de T_W^I en ω en la dirección del vector $\delta_{\bar{\omega}} = \bar{\omega} - \omega$ existe y está dada por*

$$\partial T_W^I(\omega, \bar{\omega}) = \sum_{i=1}^r \bar{\omega}_i F(\mathbf{x}_i, \tilde{\gamma}_2) - T_W^I(\omega) \quad (3-32)$$

donde $\tilde{\gamma}_2 \in \Gamma_2(\omega)$.

Demostración. Sea $g(\boldsymbol{\omega}, \boldsymbol{\gamma}_2) = \sum_{i=1}^r \omega_i F(\mathbf{x}_i, \boldsymbol{\gamma}_2)$, entonces

$$T_W^I(\boldsymbol{\omega}) = \min_{\boldsymbol{\gamma}_2 \in \Gamma_2} g(\boldsymbol{\omega}, \boldsymbol{\gamma}_2). \quad (3-33)$$

Considere la función $h : [0, 1] \times \boldsymbol{\gamma}_2$ definida por

$$h(\alpha, \boldsymbol{\gamma}_2) = g(\boldsymbol{\omega} + \alpha \delta_{\bar{\boldsymbol{\omega}}}, \boldsymbol{\gamma}_2). \quad (3-34)$$

Claramente $h(\alpha, \boldsymbol{\gamma}_2)$ es una función continua en $\alpha \in [0, 1]$ y debido a que las componentes de $\boldsymbol{\Sigma}_2(\mathbf{x}, \boldsymbol{\gamma})$ son continuas con respecto a $\boldsymbol{\gamma}_2$, se tiene que $h(\alpha, \boldsymbol{\gamma}_2)$ también es una función continua en Γ_2 . Adicionalmente, $\frac{\partial h(\alpha, \boldsymbol{\gamma}_2)}{\partial \alpha}$ existe y es continua en $[0, 1] \times \boldsymbol{\gamma}_2$. Así, aplicando el teorema A.3.4 se obtiene que la derivada direccional de T_W^I en el punto $\boldsymbol{\omega}$ en la dirección de $\delta_{\bar{\boldsymbol{\omega}}}$ existe y está dada por

$$\partial T_W^I(\boldsymbol{\omega}, \bar{\boldsymbol{\omega}}) = \min_{\boldsymbol{\gamma}_2 \in \Gamma_2(\boldsymbol{\omega})} \partial g(\boldsymbol{\omega}, \boldsymbol{\gamma}_2, \bar{\boldsymbol{\omega}}),$$

donde $\partial g(\boldsymbol{\omega}, \boldsymbol{\gamma}_2, \bar{\boldsymbol{\omega}})$ es la derivada direccional de g en $\boldsymbol{\omega}$ en la dirección de $\delta_{\bar{\boldsymbol{\omega}}}$.

Como $\Gamma_2(\boldsymbol{\omega}) = \{\tilde{\boldsymbol{\gamma}}_2\}$ debido a que $\boldsymbol{\omega}$ es un diseño T_W^I -regular, se tiene que

$$\partial T_W^I(\boldsymbol{\omega}, \bar{\boldsymbol{\omega}}) = \partial g(\boldsymbol{\omega}, \tilde{\boldsymbol{\gamma}}_2, \bar{\boldsymbol{\omega}}). \quad (3-35)$$

Luego, aplicando la definición A.3.4 de la derivada direccional se obtiene

$$\begin{aligned} \partial g(\boldsymbol{\omega}, \tilde{\boldsymbol{\gamma}}_2, \bar{\boldsymbol{\omega}}) &= \lim_{\alpha \rightarrow 0^+} \frac{g(\boldsymbol{\omega} + \alpha \delta_{\bar{\boldsymbol{\omega}}}, \tilde{\boldsymbol{\gamma}}_2) - g(\boldsymbol{\omega}, \tilde{\boldsymbol{\gamma}}_2)}{\alpha} \\ &= \lim_{\alpha \rightarrow 0^+} \frac{g((1 - \alpha)\boldsymbol{\omega} + \alpha \bar{\boldsymbol{\omega}}, \tilde{\boldsymbol{\gamma}}_2) - g(\boldsymbol{\omega}, \tilde{\boldsymbol{\gamma}}_2)}{\alpha} \\ &= \sum_{i=1}^r \bar{\omega}_i F(\mathbf{x}_i, \tilde{\boldsymbol{\gamma}}_2) - \sum_{i=1}^r \omega_i F(\mathbf{x}_i, \tilde{\boldsymbol{\gamma}}_2) \\ &= \sum_{i=1}^r \bar{\omega}_i F(\mathbf{x}_i, \tilde{\boldsymbol{\gamma}}_2) - T_W^I(\boldsymbol{\omega}). \end{aligned} \quad (3-36)$$

□

Corolario 3.5.2.1. *Bajo las condiciones del Teorema 3.5.2, si $\bar{\boldsymbol{\omega}} = \boldsymbol{\omega}_x$ es el diseño que tiene su peso concentrado en el punto \mathbf{x} entonces*

$$\partial T_W^I(\boldsymbol{\omega}, \boldsymbol{\omega}_x) = F(\mathbf{x}, \tilde{\boldsymbol{\gamma}}_2) - T_W^I(\boldsymbol{\omega}). \quad (3-37)$$

El siguiente teorema proporciona condiciones precisas para verificar si un diseño particular es T_W^I -óptimo. El resultado es una generalización del teorema de equivalencia para T-optimalidad.

Teorema 3.5.3. (Teorema de equivalencia para T_W^I -óptimalidad). Sea ω^* un diseño T_W^I -regular. Entonces, bajo los supuestos (A1)-(A3) se tiene que:

(i) Una condición necesaria y suficiente para que el diseño ω^* sea T_W^I -óptimo está dada por

$$\phi_{T_W^I}(\mathbf{x}, \omega^*) \leq 0, \quad \forall \mathbf{x} \in \mathcal{X}^n, \quad \gamma_2^* \in \Gamma_2(\omega^*) \quad (3-38)$$

donde $\phi_{T_W^I}(\mathbf{x}, \omega^*)$ es la derivada direccional en ω^* en la dirección de $\delta_{\omega_{\mathbf{x}}} = \omega_{\mathbf{x}} - \omega^*$.

(ii) La función $\phi_{T_W^I}(\mathbf{x}, \omega^*)$ alcanza su valor máximo en los puntos de soporte del diseño ω^* .

(iii) El conjunto de diseños T_W^I -óptimos es convexo.

Demostración.

(i) A partir de los teoremas 3.5.1 y 3.5.2 se sigue que T_W^I es una función cóncava cuya derivada direccional $\partial T_W^I(\omega^*, \omega)$ en el punto ω^* existe para cualquier dirección y está dada por

$$\partial T_W^I(\omega^*, \omega) = \sum_{i=1}^r \bar{\omega}_i \phi(\mathbf{x}_i, \gamma_2^*) \quad (3-39)$$

donde $\phi(\mathbf{x}, \gamma_2^*) = F(\mathbf{x}, \gamma_2^*) - T_W^I(\omega^*)$ y $\gamma_2^* \in \gamma_2(\omega^*)$. Por el corolario 3.5.2.1, $\phi(\mathbf{x}, \gamma_2^*)$ es la derivada direccional en ω^* en la dirección de $\delta_{\omega_{\mathbf{x}}}$.

Luego, aplicando el teorema A.3.3 se obtiene que ω^* es T_W^I -óptimo si y sólo si

$$\partial T_W^I(\omega^*, \omega) \leq 0, \quad \forall \omega \in \Xi. \quad (3-40)$$

Así, una condición necesaria y suficiente para que el diseño ω^* sea T_W^I -óptimo está dada por

$$\max_{\omega \in \Xi} \left\{ \sum_{i=1}^r \bar{\omega}_i \phi(\mathbf{x}_i, \gamma_2^*) \right\} \leq 0. \quad (3-41)$$

Como $\{\omega_{\mathbf{x}} : \mathbf{x} \in \bar{\mathcal{X}}^n\} \subset \Xi$ entonces

$$\max_{\mathbf{x} \in \bar{\mathcal{X}}^n} \phi(\mathbf{x}, \gamma_2^*) = \max_{\mathbf{x} \in \bar{\mathcal{X}}^n} \partial T_W^I(\omega^*, \omega_{\mathbf{x}}) \leq \max_{\omega \in \Xi} \partial T_W^I(\omega^*, \omega) = \max_{\omega \in \Xi} \sum_{i=1}^r \bar{\omega}_i \phi(\mathbf{x}_i, \gamma_2^*) \quad (3-42)$$

Además

$$\begin{aligned}
\max_{\omega \in \Xi} \sum_{i=1}^r \omega_i \phi(\mathbf{x}_i, \gamma_2^*) &= \max_{\omega \in \Xi} \sum_{i=1}^r \omega_i \partial T_W^I(\omega^*, \omega_{\mathbf{x}_i}) \\
&\leq \max_{\omega \in \Xi} \sum_{i=1}^r \omega_i \max_{\mathbf{x} \in \bar{\mathcal{X}}^n} \partial T_W^I(\omega^*, \omega_{\mathbf{x}}) \\
&= \max_{\omega \in \Xi} \sum_{i=1}^r \omega_i \max_{\mathbf{x} \in \bar{\mathcal{X}}^n} \phi(\mathbf{x}, \gamma_2^*) \\
&= \max_{\mathbf{x} \in \bar{\mathcal{X}}^n} \phi(\mathbf{x}, \gamma_2^*). \tag{3-43}
\end{aligned}$$

Luego de (3-42) y (3-43), se tiene que

$$\max_{\omega \in \Xi} \sum_{i=1}^r \omega_i \phi(\mathbf{x}_i, \gamma_2^*) = \max_{\mathbf{x} \in \bar{\mathcal{X}}^n} \phi(\mathbf{x}, \gamma_2^*). \tag{3-44}$$

Así la desigualdad en (3-41) es equivalente a

$$\max_{\mathbf{x} \in \bar{\mathcal{X}}^n} \phi(\mathbf{x}_i, \gamma_2^*) \leq 0 \tag{3-45}$$

y por tanto

$$\phi(\mathbf{x}_i, \gamma_2^*) \leq 0, \quad \forall \mathbf{x} \in \bar{\mathcal{X}}^n. \tag{3-46}$$

(ii) Suponga lo contrario, es decir, que existe un conjunto $\{\mathbf{x}_1^*, \dots, \mathbf{x}_{s_1}^*\} \subset \text{Supp}(\omega^*)$ y un escalar a tal que

$$\sum_{i=1}^{s_1} \omega_i^* \phi(\mathbf{x}_i^*, \omega^*) \leq a < 0 \quad \text{y} \quad \phi(\mathbf{x}^*, \omega^*) = 0 \quad \text{para} \quad \mathbf{x}^* \in \text{Supp}(\omega^*) \setminus \{\mathbf{x}_1, \dots, \mathbf{x}_{s_1}\}.$$

Entonces

$$\sum_{i=1}^s \omega_i^* \phi(\mathbf{x}_i^*, \omega^*) \leq a < 0, \tag{3-47}$$

donde s es el número de elementos de $\text{Supp}(\omega^*)$. Pero, a partir de (3-39) tomando $\omega = \omega^*$, se tiene que

$$\sum_{i=1}^s \omega_i^* \phi(\mathbf{x}_i^*, \omega^*) = 0, \tag{3-48}$$

lo cual es una contradicción. Por tanto la función $\phi_{T_W^I}(\mathbf{x}, \omega^*)$ alcanza su valor máximo en los puntos de soporte del diseño ω^* .

(iii) Suponga que ω_1^* y ω_2^* son T_W^I -óptimos. Entonces

$$T_W^I(\omega_1^*) = T_W^I(\omega_2^*) = \max_{\omega \in \Xi} T_W^I(\omega). \quad (3-49)$$

Sea $\omega^* = (1 - \alpha)\omega_1^* + \alpha\omega_2^*$ donde $\alpha \in [0, 1]$. A partir de la concavidad de T_W^I se obtiene que

$$\begin{aligned} T_W^I(\omega^*) &\geq (1 - \alpha)T_W^I(\omega_1^*) + \alpha T_W^I(\omega_2^*) \\ &= \max_{\omega \in \Xi} T_W^I(\omega). \end{aligned} \quad (3-50)$$

Por tanto

$$T_W^I(\omega) \leq T_W^I(\omega^*), \quad \forall \omega \in \Xi. \quad (3-51)$$

□

3.6. Ejemplos

En esta sección se presentan dos ejemplos que ilustran la metodología propuesta. En el primer ejemplo se considera un modelo polinomial de decrecimiento exponencial utilizado para describir datos de experimentos que se obtienen en un periodo de tiempo a partir de una perturbación bioquímica en un organismo. En el segundo ejemplo se considera un modelo compartimental utilizado con frecuencia en experimentos farmacocinéticos para describir la concentración de un medicamento en función del tiempo. Para los dos ejemplos el objetivo es hallar diseños que permitan discriminar entre dos modelos para la variabilidad intra-sujeto. Con el fin de mostrar las ventajas del criterio propuesto, se evalúa la eficiencia de cuatro diseños utilizados comúnmente con respecto al diseño propuesto.

3.6.1. Modelo polinomial de decrecimiento exponencial

Modelos

En Hand y Crowder (1996) se describe un estudio bioquímico realizado en dos momentos con el fin de evaluar el efecto de un tipo de aditivo alimentario en ciertos compuestos de la sangre. Los datos del primer experimento corresponden a los niveles de glucosa en la sangre de siete individuos medidos durante un periodo de tiempo después de ingerir una bebida alcohólica. Por cada individuo se tomaron 14 muestras de sangre sobre un periodo de 0 a 5 horas. Los datos son tomados de Hand y Crowder (1996) y se encuentran en el marco de datos `Glucose2` de la librería `nlme` del software R. En lo que sigue por simplicidad en los ajustes el individuo 6 no es incluido en el análisis.

Los datos se representan gráficamente en la Figura 3-1.

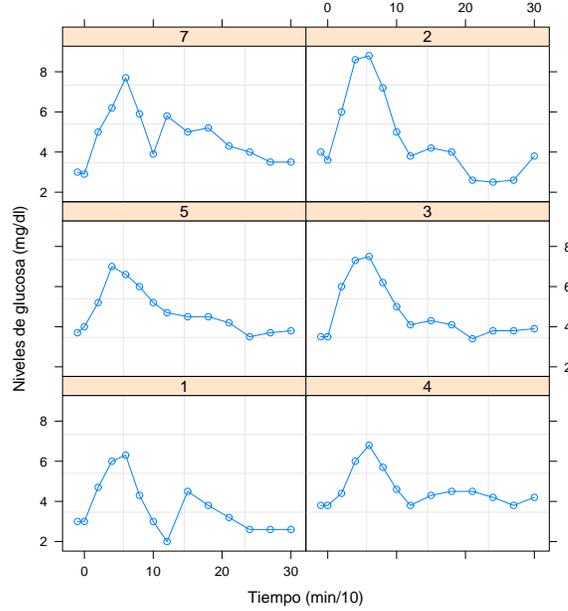


Figura 3-1.: Niveles de glucosa para seis individuos.

Crowder y Tredger (1981) sugirieron una clase de modelos de regresión polinomiales de decrecimiento exponencial que se pueden utilizar para describir datos de experimentos que se obtienen en un periodo de tiempo a partir de una perturbación bioquímica en un organismo. Una curva polinomial de decrecimiento exponencial está dada por

$$f(x) = \beta_1 + \beta_2(x)x^q \exp(-\beta_3x), \quad \forall x \geq 0 \quad (3-52)$$

donde x representa el tiempo, $\beta_2(x) = \beta_2 + \beta_{2(1)}x + \dots + \beta_{2(p)}x^p$; p, q son números enteros conocidos y $(\beta_1, \beta_2, \beta_{2(1)}, \dots, \beta_{2(p)}, \beta_3)^T$ es el vector de parámetros a estimar. Se supone que β_2 es un valor diferente de cero, $\beta_3 > 0$ y $q \geq 0$. Las características más importantes que se reflejan en este tipo de curvas son: (i) la intensidad del decrecimiento representada por el parámetro β_3 , (ii) el número y naturaleza de los puntos de inflexión y (iii) el comportamiento en $x = 0$ y $x = \infty$. Si $q > 0$, $f(\infty) = f(0)$ entonces el proceso regresa eventualmente a su nivel inicial y en cualquier otro caso si $q = 0$, $f(\infty) = \alpha$ y $f(0) = \beta_1 + \beta_2$.

Una representación alternativa es la siguiente:

$$f(x) = \beta_1 + \beta_2\gamma(x)x^q \exp(-\beta_3x), \quad (3-53)$$

donde $\gamma(x) = 1 + \gamma_1x + \dots + \gamma_px^p$ y $\gamma_i = \beta_{2(i)}/\beta_2$. De esta forma, β_1 y β_2 se identifican como parámetros de localización y escala, mientras que $(\gamma_1, \dots, \gamma_p)^T$ determinan la forma de la curva.

Para el estudio descrito previamente, se considera el siguiente modelo marginal de decrecimiento exponencial con $p = 0$ y $q = 3$:

$$y_{il} = \beta_1 + \beta_2x_{il}^3 \exp(-\beta_3x_{il}) + b_i + \varepsilon_{il}, \quad (3-54)$$

donde b_i es el efecto aleatorio asociado al i -ésimo individuo y ε_{il} es el término del error aleatorio intra-individuo. Sea $\mathbf{b} = (b_1, \dots, b_N)^T$ y $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in})^T$. Asuma que:

$$\mathbf{b} \sim \mathcal{N}_N(\mathbf{0}, \psi \mathbf{I}_N), \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})), \quad \text{Cov}(\mathbf{b}, \boldsymbol{\varepsilon}_i) = \mathbf{0}. \quad (3-55)$$

Para ilustrar el criterio de optimalidad propuesto, se consideran dos modelos alternativos para la variabilidad intra-individual. En el primero se asume que los errores aleatorios intra-individuo tienen varianza constante y estructura de correlación dada por un modelo autorregresivo AR(1) continuo con el tiempo como covariable. Esta estructura fue utilizada por Hand y Crowder (1996) quienes propusieron para estos datos ajustar un modelo mixto con efectos aleatorios en los parámetros β_1 y β_2 . En el segundo modelo se asume la estructura usual: errores independientes con varianza constante. En forma compacta, estos modelos se pueden escribir como:

\mathcal{M}_1 . Modelo AR(1) continuo y varianza constante

$$\begin{aligned} \boldsymbol{\varepsilon}_i &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})) \\ \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) &= \sigma^2 \mathbf{W}(\mathbf{x}_i, \rho) \end{aligned}$$

donde $\mathbf{W}(\mathbf{x}_i, \rho) = \{w_{sk}(\mathbf{x}_i, \rho)\}_{s,k=1}^n$, $w_{sk}(\mathbf{x}_i, \rho) = \rho^{|x_{is} - x_{ik}|}$, ($-1 < \rho < 1$).

\mathcal{M}_2 . Errores no correlacionados con varianza constante

$$\begin{aligned} \boldsymbol{\varepsilon}_i &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})) \\ \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) &= \sigma^2 \mathbf{I}_n. \end{aligned}$$

Los dos modelos fueron ajustados a los datos usando la función `nlme` del software estadístico R. Los parámetros estimados se muestran a continuación:

$$\begin{aligned} \mathcal{M}_1 &: \widehat{\boldsymbol{\beta}}_1 = (3.76, 0.43, 0.57)^T, \quad \widehat{\psi}_1 = 0.20^2, \quad \widehat{\sigma}_1^2 = 0.77^2, \quad \widehat{\rho} = 0.72 \\ \mathcal{M}_2 &: \widehat{\boldsymbol{\beta}}_2 = (3.77, 0.53, 0.60)^T, \quad \widehat{\psi}_2 = 0.34^2, \quad \widehat{\sigma}_2^2 = 0.71^2. \end{aligned} \quad (3-56)$$

En las gráficas **3-2** y **3-3** se muestra las curvas individuales ajustadas para cada modelo. Se observa que los dos modelos son adecuados para representar el perfil de los niveles de glucosa. Para la construcción del diseño, los vectores $\widehat{\boldsymbol{\beta}}_1$ y $\widehat{\boldsymbol{\beta}}_2$ son los vectores de parámetros locales.

Diseño T_W^I -óptimo

Debido a que el modelo \mathcal{M}_2 está anidado dentro de \mathcal{M}_1 , se asume que \mathcal{M}_1 es el modelo correcto con vector de parámetros dado por

$$\boldsymbol{\theta}_1^0 = (\boldsymbol{\beta}_1^T, \sigma_1^2, \psi_1, \lambda)^T$$

donde $\boldsymbol{\beta}_1 = (3.76, 0.43, 0.57)^T$, $\sigma_1^2 = 0.77^2$, $\psi_1 = 0.20^2$ y $\lambda = \rho = 0.72$.

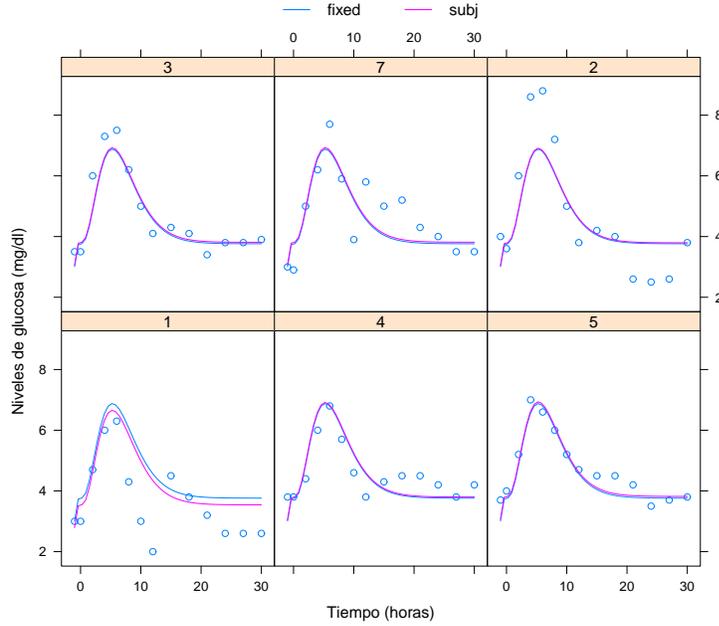


Figura 3-2.: Ajuste del Modelo 1.

Para el modelo alternativo \mathcal{M}_2 el vector de parámetros poblacionales es

$$\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2^T, \sigma_2^2, \psi_2, 0)^T. \quad (3-57)$$

Con fines ilustrativos, se tomó a \mathcal{X} como el conjunto de tiempos de medición del experimento sin incluir $t = -1$

$$\mathcal{X} = \{0, 2, 4, 6, 8, 10, 12, 15, 18, 21, 24, 27, 30\} \text{ min}/10. \quad (3-58)$$

y se fijaron tres muestras por individuo, es decir, $n = 3$. Así, la región de diseño $\bar{\mathcal{X}}^3$ contiene $C_{13,3} = 286$ combinaciones posibles de secuencias de elementos tomados de \mathcal{X} :

$$\bar{\mathcal{X}}^3 = \{(x_1, x_2, x_3) : x_s \in \mathcal{X}\}. \quad (3-59)$$

Para hallar el diseño T_W^I -óptimo local se usaron los valores de los estimadores de los parámetros como parámetros locales y la región de diseño $\bar{\mathcal{X}}^3$. El diseño óptimo se calculó optimizando el criterio T_W^I a través de un algoritmo implementado en el software estadístico R [41]. La función `nlimb` fue utilizada en la optimización sobre la región $\bar{\mathcal{X}}^3$. El código fuente en R para el cálculo de los diseños puede verse en el Apéndice B.

El diseño poblacional resultante es un diseño de seis puntos dado por

$$\zeta_{T_W^I}^* = \left\{ \begin{array}{cccccc} (0, 2, 4) & (0, 2, 18) & (0, 2, 21) & (0, 2, 24) & (0, 2, 27) & (0, 2, 30) \\ 0.26 & 0.14 & 0.14 & 0.17 & 0.19 & 0.10 \end{array} \right\} \quad (3-60)$$

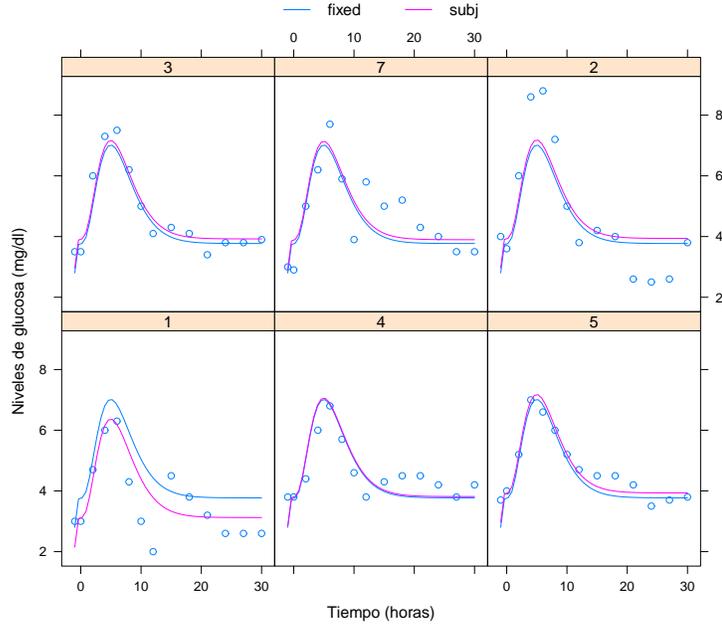


Figura 3-3.: Ajuste del Modelo 2.

y el vector de parámetros γ_2^* del modelo \mathcal{M}_2 obtenido en el proceso de optimización es

$$\begin{aligned}\gamma_2^* &= (\sigma_2^{2*}, \psi_2^*)^T \\ &= (0.72^2, 0.41^2)^T\end{aligned}$$

Una interpretación práctica del diseño es la siguiente: si el experimento se repitiera fijando tres muestras por individuo y el objetivo es obtener observaciones que permitan discriminar entre los modelos \mathcal{M}_1 y \mathcal{M}_2 , entonces para un tamaño de muestra fijado N : el 26 % de los pacientes se deben medir en 0,2 y 4 (min/10) después de ingerir la bebida, el 14 % de los pacientes en (0,2,18) (min/10) y así sucesivamente. De acuerdo al diseño obtenido todos los pacientes se deben medir en 0 y 2 min, es decir, inmediatamente después de la ingesta de la bebida, y la tercera muestra se toma en un lapso de tiempo mayor excepto para el primer grupo.

Con el teorema de equivalencia, se puede verificar gráficamente que en efecto el diseño obtenido es T_W^I -óptimo local. Primero se enumeran todas las secuencias de tiempos de medición en $\bar{\mathcal{X}}^3$ en el siguiente orden:

$$\mathbf{x}_1 = (0, 2, 4), \mathbf{x}_2 = (0, 2, 6), \dots, \mathbf{x}_{285} = (21, 27, 30), \mathbf{x}_{286} = (24, 27, 30).$$

y se calcula la función de sensibilidad $F(\mathbf{x}_k, \gamma_2^*) - T_W^I(\omega^*)$ para todas las 283 secuencias. Luego se grafica dicha función como una función del índice k , ver Figura 3-4. Se observa que para cualquier $\mathbf{x} \in \bar{\mathcal{X}}^3$

$$F(\mathbf{x}, \gamma_2^*) - T_W^I(\omega^*) \leq 0$$

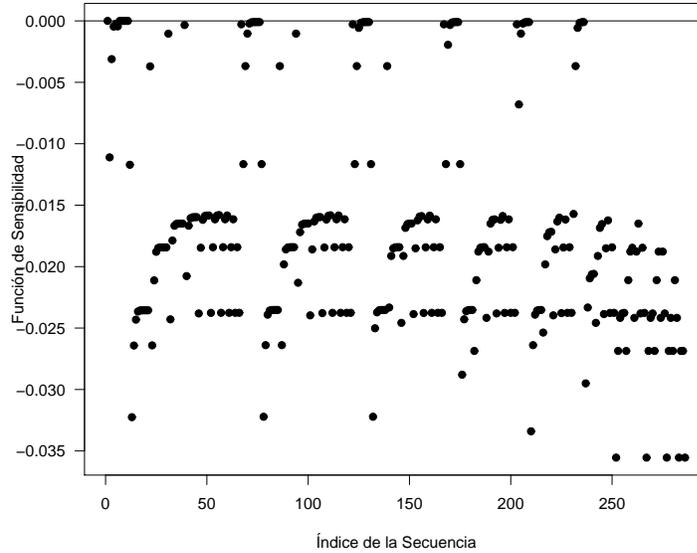


Figura 3-4.: Función de sensibilidad para los vectores de muestreo.

y en los puntos de soporte del diseño la función de sensibilidad es cero. Además, se observa que existen muchos diseños que también son buenos.

Comparación con otros diseños

Con el fin de demostrar las ventajas de la metodología propuesta, se comparó el rendimiento de diferentes diseños poblacionales comúnmente utilizados en la práctica con el diseño T_W^I -óptimo, mediante el cálculo de la T_W^I -eficiencia. Estos diseños se describen a continuación.

- (i) El diseño producto de D-optimalidad sugerido por Waterhouse *et al.* (2005). El criterio producto está basado en el producto de los criterios de D-optimalidad para cada uno de los modelos \mathcal{M}_1 y \mathcal{M}_2 . Específicamente, el diseño producto para dos modelos está definido por

$$\boldsymbol{\omega}^* = \arg \max_{\boldsymbol{\omega} \in \Xi} |M_1(\boldsymbol{\omega}, \boldsymbol{\theta}_1)|^{1/v_1} |M_2(\boldsymbol{\omega}, \boldsymbol{\theta}_2)|^{1/v_2}, \quad (3-61)$$

donde $M_j(\boldsymbol{\omega}, \boldsymbol{\theta}_j)$ es la matriz de información de Fisher para el j -ésimo modelo con vector de parámetros $\boldsymbol{\theta}_j$ y $v_j = \dim(\boldsymbol{\theta}_j)$. En el capítulo 5, se presenta una expresión para la matriz de información y algunas propiedades de los diseños D-óptimos para modelos de efectos mixtos.

- (ii) Los diseños D-óptimos para cada modelo:

$$\boldsymbol{\omega}_{D_1}^* = \arg \max_{\boldsymbol{\omega} \in \Xi} \log \det M(\boldsymbol{\omega}, \boldsymbol{\theta}_1) \quad (3-62)$$

y

$$\boldsymbol{\omega}_{D_2}^* = \arg \max_{\boldsymbol{\omega} \in \Xi} \log \det M(\boldsymbol{\omega}, \boldsymbol{\theta}_2). \quad (3-63)$$

- (iii) Un diseño de rejilla obtenido a partir de la división del conjunto de tiempos \mathcal{X} en cuatro subconjuntos. Los diseños de rejilla sugeridos por Fedorov y Leonov (2007) corresponden a secuencias alternativas de puntos de medición que se obtienen dividiendo el conjunto $\mathcal{X} = \{x_1, \dots, x_s\}$ (la rejilla inicial) en dos o más subconjuntos. Estos autores mostraron la utilidad de esta clase de diseños en la estimación de ciertas funciones de interés en estudios farmacocinéticos poblacionales. Para comparar el diseño obtenido bajo el criterio T_W^I , se consideró un diseño de rejilla de cuarto orden con secuencias $\{x_1, x_5, x_9\}$, $\{x_2, x_6, x_{10}\}$, $\{x_3, x_7, x_{11}\}$ y $\{x_4, x_8, x_{12}\}$, y la misma proporción de individuos en cada una de ellas.

Los diseños anteriores se denotan por ζ_P^* , $\zeta_{D_1}^*$, $\zeta_{D_2}^*$ y ζ_R , respectivamente. Para calcular la eficiencia de un diseño ζ con respecto al criterio T_W^I se utiliza la razón

$$T_W^I - \text{ef} = \frac{T_W^I(\zeta)}{T_W^I(\zeta_{T_W^I}^*)} \quad (3-64)$$

donde $\zeta_{T_W^I}^*$ es el diseño T_W^I -óptimo. En la Tabla **3-1** se presentan los diseños poblacionales obtenidos junto con el cálculo de la T_W^I -eficiencia.

Tabla 3-1.: T_W^I -eficiencias de diseños poblacionales

Diseño	$T_W^I - \text{ef}(\zeta)$
$\zeta_{T_W^I}^* = \left\{ \begin{array}{cccccc} (0, 2, 4) & (0, 2, 18) & (0, 2, 21) & (0, 2, 24) & (0, 2, 27) & (0, 2, 30) \\ 0.26 & 0.14 & 0.14 & 0.17 & 0.19 & 0.10 \end{array} \right\}$	1
$\zeta_P^* = \left\{ \begin{array}{cccc} (0, 4, 6) & (4, 6, 8) & (4, 6, 30) & (6, 8, 30) \\ 0.18 & 0.10 & 0.50 & 0.22 \end{array} \right\}$	0.9455
$\zeta_{D_1}^* = \left\{ \begin{array}{ccccccc} (0, 2, 4) & (0, 2, 8) & (0, 4, 8) & (0, 8, 27) & (0, 8, 30) & (4, 6, 8) & (4, 8, 30) \\ 0.27 & 0.08 & 0.09 & 0.03 & 0.06 & 0.28 & 0.19 \end{array} \right\}$	0.7995
$\zeta_{D_2}^* = \left\{ \begin{array}{ccc} (0, 4, 6) & (0, 6, 8) & (0, 6, 30) \\ 0.73 & 0.21 & 0.06 \end{array} \right\}$	0.7653
$\zeta_R = \left\{ \begin{array}{cccc} (0, 8, 18) & (2, 10, 21) & (4, 12, 24) & (6, 15, 27) \\ 0.25 & 0.25 & 0.25 & 0.25 \end{array} \right\}$	0.0001

Se observa que los diseños $\zeta_{D_1}^*$ y $\zeta_{D_2}^*$ tienen eficiencias similares del 80% y 77% respectivamente. En este caso, el diseño ζ_P^* tiene una eficiencia comparable con respecto al diseño T_W^I -óptimo mientras que el diseño de rejilla no es eficiente para discriminación.

3.6.2. Modelo de dos compartimientos

En este ejemplo se consideran los datos de un estudio realizado sobre la farmacocinética del medicamento indometacina (ver Davidian y Giltinan, 1995), un antiinflamatorio que reduce el dolor y la inflamación. A seis individuos voluntarios se les administró por vía intravenosa una dosis igual de indometacina y se midieron sus concentraciones en plasma 11 veces en un periodo de 15 minutos a 8 horas después de la inyección. Los datos son tomados de Davidian y Giltinan (1995) y se encuentran en el marco de datos `Indometh` de la librería `nlme` del software R.

Estos datos se representan gráficamente en la Figura 3-5. Del gráfico se observa que aunque

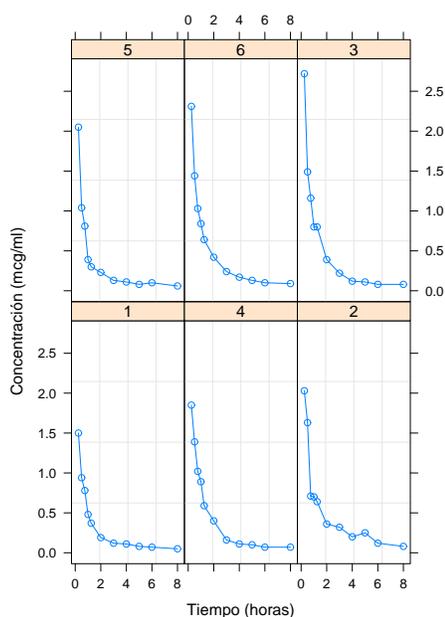


Figura 3-5.: Concentración de Indometacina para los seis individuos.

la forma de las curvas de concentración son similares éstas difieren para cada individuo. Un enfoque usual para modelar datos que provienen de un experimento farmacocinético es representar el organismo como un sistema de compartimientos en los cuales la transferencia del medicamento sigue una cinética de primer orden, es decir, las tasas de transferencia de un compartimiento a otro son proporcionales a la cantidad de sustancia en el compartimiento proveedor. En esta clase de modelos, la concentración del medicamento en cada compartimiento a través del tiempo está determinada por un sistema lineal de ecuaciones diferenciales cuya solución se puede expresar como una combinación lineal de términos exponenciales. Por lo general, muchos medicamentos dan al organismo las características de un modelo de dos compartimientos. Bajo esta representación, luego de una inyección intravenosa, la concentración del medicamento en el compartimiento central (plasma o sistema circulatorio) a lo

largo del tiempo está dada por

$$f(x) = \beta_1 \exp(-\beta_2 x) + \beta_3 \exp(-\beta_4 x), \quad \beta_1, \dots, \beta_4 > 0. \quad (3-65)$$

donde x es el tiempo, $(\beta_1, \beta_2, \beta_3, \beta_4)^T$ es un vector de parámetros con $\beta_2 > \beta_4$. El primer término del modelo determina la fase inicial del medicamento (decrecimiento en la curva). La fase de eliminación está determinada principalmente por el segundo término. Para el estudio de la indometacina, se considera el siguiente modelo marginal:

$$y_{il} = \beta_1 \exp(-\beta_2 x_{il}) + \beta_3 \exp(-\beta_4 x_{il}) + b_i + \varepsilon_{il}, \quad (3-66)$$

donde b_i es el efecto aleatorio asociado al i -ésimo individuo y ε_{il} es el término del error aleatorio intra-individuo. Sea $\mathbf{b} = (b_1, \dots, b_N)^T$ y $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in})^T$. Asuma que:

$$\mathbf{b} \sim \mathcal{N}_N(\mathbf{0}, \psi \mathbf{I}_N), \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})), \quad Cov(\mathbf{b}, \boldsymbol{\varepsilon}_i) = \mathbf{0}. \quad (3-67)$$

En este caso

$$\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) = (\beta_1 \exp(-\beta_2 x_{i1}) + \beta_3 \exp(-\beta_4 x_{i1}), \dots, \beta_1 \exp(-\beta_2 x_{in}) + \beta_3 \exp(-\beta_4 x_{in}))^T \quad (3-68)$$

y $\mathbf{Z}_i = \mathbf{1}_n$.

Las dos estructuras consideradas para la variabilidad intra-individual son: (i) la estructura utilizada por Lin y Wei (2007): un modelo de varianza con función exponencial y un modelo autorregresivo AR(1) continuo, y (ii) la estructura usual: errores independientes con varianza constante. En forma compacta, estos modelos se pueden escribir como:

\mathcal{M}_1 . *Función de varianza exponencial y modelo AR(1) continuo*

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})), \quad \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = \mathbf{G}^{1/2}(\mathbf{x}_i, \sigma^2, \alpha) \mathbf{W}(\mathbf{x}_i, \rho) \mathbf{G}^{1/2}(\mathbf{x}_i, \sigma^2, \alpha)$$

donde

$$\begin{aligned} \mathbf{G}(\mathbf{x}_i, \sigma^2, \alpha) &= \sigma^2 \text{diag}(\exp(-\alpha x_{i1}), \dots, \exp(-\alpha x_{in})) \\ \mathbf{W}(\mathbf{x}_i, \rho) &= \{w_{sk}(\mathbf{x}_i, \rho)\}_{s,k=1}^n, \quad w_{sk}(\mathbf{x}_i, \rho) = \rho^{|x_{is} - x_{ik}|}, \quad (-1 < \rho < 1). \end{aligned}$$

\mathcal{M}_2 . *Errores no correlacionados con varianza constante*

$$\begin{aligned} \boldsymbol{\varepsilon}_i &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})) \\ \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) &= \sigma^2 \mathbf{I}_n. \end{aligned}$$

Los vectores de parámetros estimados de los dos modelos se muestran a continuación.

$$\begin{aligned} \mathcal{M}_1 &: \widehat{\boldsymbol{\beta}}_1 = (2.92, 1.92, 0.28, 0.17)^T, \quad \widehat{\psi}_1 = 0.012^2, \quad \widehat{\sigma}_1^2 = 0.36^2, \quad \widehat{\rho}_1 = 0.76, \quad \widehat{\alpha}_1 = 0.48 \\ \mathcal{M}_2 &: \widehat{\boldsymbol{\beta}}_2 = (2.76, 2.48, 0.63, 0.37)^T, \quad \widehat{\psi}_2 = 0.097^2, \quad \widehat{\sigma}_2^2 = 0.14^2. \end{aligned}$$

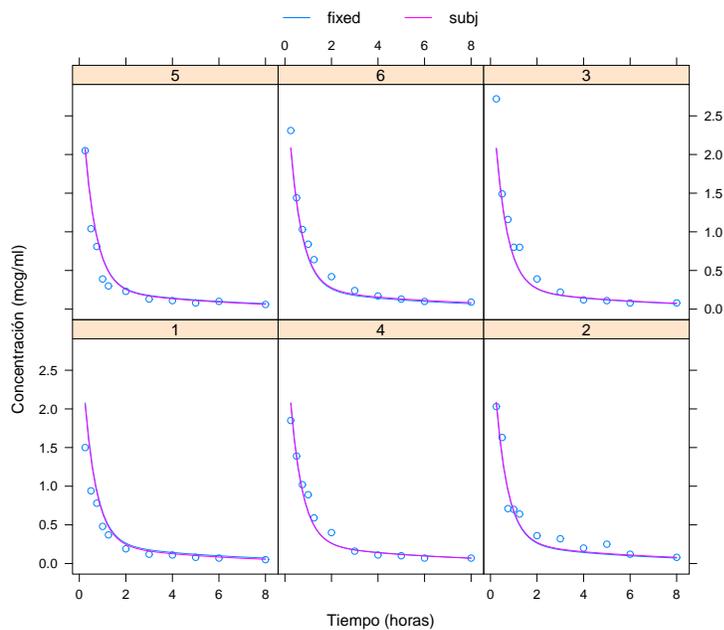


Figura 3-6.: Ajuste del Modelo 1.

En las gráficas 3-6 y 3-7 se muestra las curvas individuales ajustadas para cada modelo. Se observa que el modelo 1 es el más adecuado para representar el perfil de la concentración de indometacina.

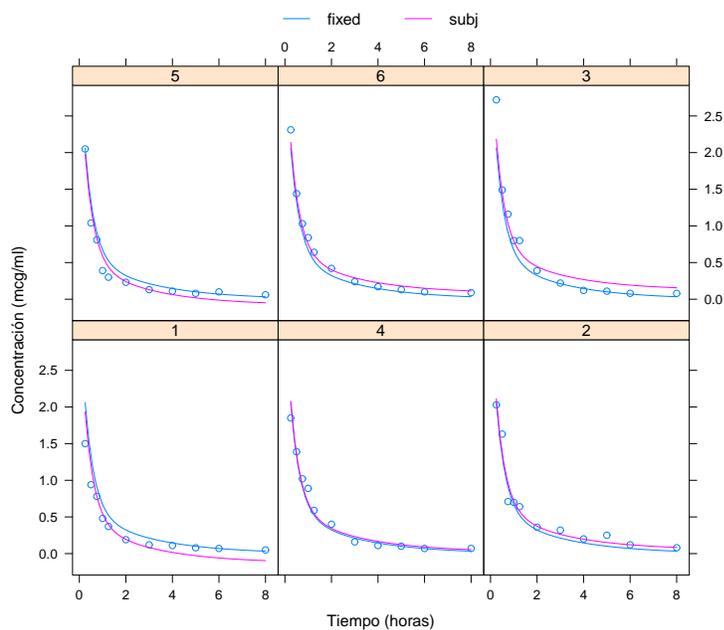


Figura 3-7.: Ajuste del Modelo 2.

Diseño T_W^I -óptimo

Debido a que el modelo \mathcal{M}_2 está anidado dentro de \mathcal{M}_1 , se asume que \mathcal{M}_1 es el modelo correcto con vector de parámetros $\boldsymbol{\gamma}_1$ dado por

$$\boldsymbol{\gamma}_1^0 = (\sigma_1^2, \psi_1, \boldsymbol{\lambda}^T)^T$$

donde $\sigma_1^2 = 0.36^2$, $\psi_1 = 0.012^2$ y $\boldsymbol{\lambda} = (0.48, 0.76)^T$.

Para el modelo alternativo \mathcal{M}_2 el vector de parámetros poblacionales $\boldsymbol{\gamma}_2$ es

$$\boldsymbol{\gamma}_2 = (\sigma_2^2, \psi_2, \mathbf{0}^T)^T. \quad (3-69)$$

Con fines ilustrativos, se tomó a \mathcal{X} como el conjunto de tiempos de medición del experimento

$$\mathcal{X} = \{0.25, 0.50, 0.75, 1.00, 1.25, 2.00, 3.00, 4.00, 5.00, 6.00, 8.00\} \text{ horas} \quad (3-70)$$

y se fijaron cuatro muestras por individuo, es decir, $n = 4$. Así, la región de diseño $\bar{\mathcal{X}}^4$ contiene $C_{11,4} = 330$ combinaciones posibles de secuencias de elementos tomados de \mathcal{X} :

$$\bar{\mathcal{X}}^4 = \{(x_1, x_2, x_3, x_4) : x_s \in \mathcal{X}\}. \quad (3-71)$$

Para hallar el diseño T_W^I -óptimo local se usaron los valores de los estimadores de los parámetros como parámetros locales y la región de diseño $\bar{\mathcal{X}}^4$.

El diseño poblacional resultante es un diseño de dos puntos dado por

$$\zeta_{T_W^I}^* = \left\{ \begin{array}{cc} (0.25, 0.50, 6, 8) & (4, 5, 6, 8) \\ 0.94 & 0.06 \end{array} \right\} \quad (3-72)$$

y el vector de parámetros $\boldsymbol{\gamma}_2^*$ del modelo \mathcal{M}_2 obtenido en el proceso de optimización es

$$\boldsymbol{\gamma}_2^* = (0.19^2, 0.74^2)^T.$$

De acuerdo al diseño obtenido la mayoría de los pacientes, aproximadamente un 94 %, se deben medir en 0.25, 0.5, 6 y 8 horas y un 6 % en 4,5,6 y 8 horas. En la Figura 3-8 se verifica que en efecto el diseño hallado es T_W^I -óptimo local, es decir, se muestra que para cualquier $\boldsymbol{x} \in \bar{\mathcal{X}}^4$

$$F(\boldsymbol{x}, \boldsymbol{\gamma}_2^*) - T_W^I(\boldsymbol{\omega}^*) \leq 0$$

y es cero en los puntos de soporte del diseño.

Comparación con otros diseños

En la Tabla 3-2 se muestran los diseños D-óptimos para cada modelo, el diseño producto y un diseño de rejilla de tercer orden con tres grupos asignados a las secuencias de medición (x_1, x_4, x_7, x_{10}) , (x_2, x_5, x_8, x_{11}) y (x_3, x_6, x_9) , respectivamente.

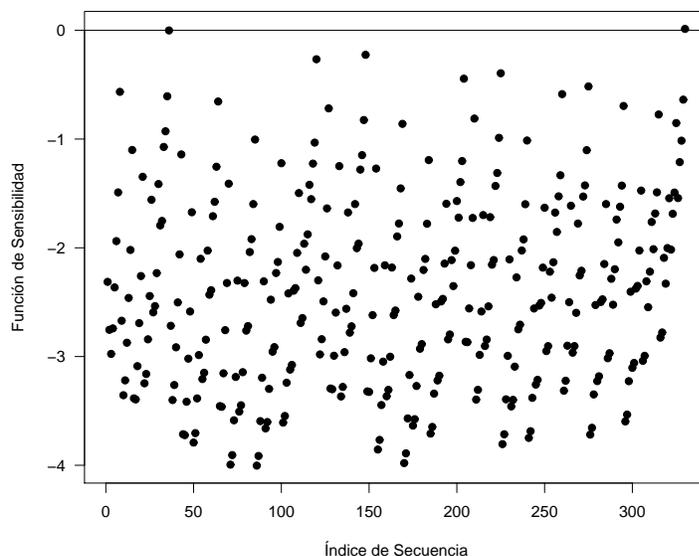


Figura 3-8.: Función de sensibilidad para T_W^I -optimalidad local.

Tabla 3-2.: T_W^I -eficiencias de diseños poblacionales

Diseño	$T_W^I - \text{ef}(\zeta)$
$\zeta_{T_W^I}^* = \left\{ \begin{array}{cc} (0.25, 0.50, 6, 8) & (4, 5, 6, 8) \\ 0.94 & 0.06 \end{array} \right\}$	1
$\zeta_P^* = \left\{ \begin{array}{cccc} (0.25, 0.50, 1.25, 4) & (0.25, 0.50, 2, 8) & (0.25, 0.75, 4, 8) & (0.25, 1.25, 4, 8) \\ 0.30 & 0.55 & 0.13 & 0.02 \end{array} \right\}$	0.6171
$\zeta_{D_1}^* = \left\{ \begin{array}{cc} (0.25, 0.50, 0.75, 2) & (0.25, 0.50, 2, 8) \\ 0.13 & 0.87 \end{array} \right\}$	0.6664
$\zeta_{D_2}^* = \left\{ \begin{array}{cc} (0.25, 0.50, 1.25, 4) & (0.25, 0.50, 2, 4) \\ 0.70 & 0.30 \end{array} \right\}$	0.3998
$\zeta_R = \left\{ \begin{array}{ccc} (0.25, 1, 3, 6) & (0.50, 1.25, 4, 8) & (0.75, 2, 5) \\ 0.33 & 0.33 & 0.33 \end{array} \right\}$	0.3643

Las T_W^I -eficiencias de estos diseños se encuentran en la Tabla **3-2**. Se observa que las eficiencias en todos los diseños son inferiores al 67%. Tanto el diseño $\zeta_{D_2}^*$ como el diseño ζ_R son poco eficientes para discriminar mientras que el diseño $\zeta_{D_1}^*$ tiene la mayor T_W^I -eficiencia. En este caso el diseño producto tiene cuatro puntos de soporte y su eficiencia es solamente del 62% cuando compite con el diseño $\zeta_{T_W^I}^*$.

4. Diseños óptimos para discriminación entre modelos marginales mixtos tipo II

En este capítulo los resultados presentados previamente en el capítulo anterior se extienden a una clase más general de modelos mixtos. Específicamente, se propone una generalización del criterio de T-optimalidad para modelos marginales tipo II. Se muestra, además, cómo se puede aplicar el criterio propuesto para hallar diseños óptimos que permitan discriminar entre dos modelos generales de efectos mixtos anidados con diferentes matrices de covarianza intra-individual.

En la sección 4.1 se presenta el modelo y se define el criterio de optimalidad para discriminación. En la sección 4.2 se establece el teorema de equivalencia para el criterio propuesto. En la sección 4.3 se considera el problema de discriminación para modelos generales de efectos mixtos. Finalmente, en la sección 4.4 se presenta un ejemplo ilustrativo.

4.1. Modelo y un criterio de optimalidad

Considere el modelo marginal (3-1) y suponga que la matriz \mathbf{Z}_i depende del vector $\boldsymbol{\beta}$ de parámetros poblacionales. Entonces el vector de observaciones \mathbf{y}_i del i -ésimo individuo es representado como

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{Z}_i(\boldsymbol{\beta})\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N. \quad (4-1)$$

Suponga además que $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}))$, entonces bajo los demás supuestos descritos en la sección 3.1, se tiene que la distribución marginal de \mathbf{y}_i es normal n -variada con vector de medias y matriz de varianzas y covarianzas:

$$E(\mathbf{y}_i) = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) \quad (4-2)$$

$$\begin{aligned} Cov(\mathbf{y}_i) &= \mathbf{Z}_i(\boldsymbol{\beta})\mathbf{D}\mathbf{Z}_i^T(\boldsymbol{\beta}) + \mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}) \\ &\equiv \Sigma(\mathbf{x}_i, \boldsymbol{\theta}) \end{aligned} \quad (4-3)$$

donde $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \boldsymbol{\psi}^T, \boldsymbol{\lambda}^T)^T$. Se observa que, en este caso, la matriz de covarianzas marginal depende del vector $\boldsymbol{\beta}$ de parámetros poblacionales. Este modelo es llamado *modelo*

marginal tipo II. Para esta clase de modelos, se pueden ajustar estructuras de covarianza intra-individuos más generales, como por ejemplo, el modelo:

$$\mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}, \sigma^2, \alpha) = \sigma^2 \text{diag}(f^{2\alpha}(x_{i1}, \boldsymbol{\beta}), \dots, f^{2\alpha}(x_{in}, \boldsymbol{\beta})). \quad (4-4)$$

Esta matriz supone errores no correlacionados con varianza proporcional a una potencia de la respuesta media. Adicionalmente, si las observaciones se miden a través del tiempo, es posible considerar un modelo para la correlación serial, como por ejemplo el modelo autorregresivo (AR) de primer orden para observaciones igualmente espaciadas.

El modelo marginal también se puede escribir como

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) + \boldsymbol{\epsilon}_i,$$

donde $\boldsymbol{\epsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \Sigma(\mathbf{x}_i, \boldsymbol{\theta}))$.

Como se procedió en el capítulo anterior, sean \mathcal{M}_1 y \mathcal{M}_2 dos modelos con estructura dada por (4-1), los cuales asumen modelos diferentes para la variabilidad intra-individual. Suponiendo que \mathcal{M}_2 está anidado dentro de \mathcal{M}_1 con respecto a la matriz de covarianzas intra-individuos, estos modelos se pueden representar como:

$$\mathcal{M}_1 : \boldsymbol{\epsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \Sigma_1(\mathbf{x}_i, \boldsymbol{\theta}_1)), \quad \boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1^T, \sigma_1^2, \boldsymbol{\psi}_1^T, \boldsymbol{\lambda}_1^T)^T \in \Theta_1 \subset \mathbb{R}^{\nu_1}$$

$$\mathcal{M}_2 : \boldsymbol{\epsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \Sigma_2(\mathbf{x}_i, \boldsymbol{\theta}_2)), \quad \boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2^T, \sigma_2^2, \boldsymbol{\psi}_2^T, \boldsymbol{\lambda}_2^T)^T \in \Theta_2 \subset \Theta_1$$

donde $\Theta_2 = \{\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2^T, \sigma_2^2, \boldsymbol{\psi}_2^T, \boldsymbol{\lambda}_2^T)^T \in \Theta_1 : h_\tau(\boldsymbol{\lambda}_2) = 0, \tau = 1, \dots, \kappa\}$.

Con el fin de discriminar entre \mathcal{M}_1 y \mathcal{M}_2 se propone el siguiente criterio de optimalidad.

Definición 4.1.1. *Asumiendo que \mathcal{M}_1 es el modelo correcto con vector de parámetros $\boldsymbol{\theta}_1^0$ conocido, un diseño poblacional óptimo local ζ_N^* para discriminar entre \mathcal{M}_1 y \mathcal{M}_2 verifica*

$$\zeta_N^* = \arg \max_{\zeta_N} T_W^{II}(\zeta_N) \quad (4-5)$$

donde

$$T_W^{II}(\zeta_N) = \min_{\boldsymbol{\theta}_2 \in \Theta_2} \sum_{k=1}^s \frac{N_k}{N} F(\mathbf{x}_k, \boldsymbol{\theta}_2) \quad (4-6)$$

con

$$\begin{aligned} F(\mathbf{x}, \boldsymbol{\theta}_2) &= \text{tr}[\Sigma(\mathbf{x})\Sigma^{-1}(\mathbf{x}, \boldsymbol{\theta}_2)] - \log \det[\Sigma(\mathbf{x})\Sigma^{-1}(\mathbf{x}, \boldsymbol{\theta}_2)] \\ &\quad (\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}_2))^T \Sigma^{-1}(\mathbf{x}, \boldsymbol{\theta}_2) (\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}_2)) \end{aligned} \quad (4-7)$$

donde $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_1^0)$ y $\Sigma(\mathbf{x}) = \Sigma(\mathbf{x}, \boldsymbol{\theta}_1^0)$.

A continuación se presenta la justificación del criterio bajo argumentos similares a los empleados en la sección 3.4.1.

Justificación del criterio

Sea $\boldsymbol{\theta}^0$ el valor verdadero pero desconocido del parámetro $\boldsymbol{\theta}$. Suponga que se quiere probar el siguiente sistema de hipótesis:

$$H_0 : \boldsymbol{\theta}^0 \in \Theta_2 \quad \text{versus} \quad H_1 : \boldsymbol{\theta}^0 \in \Theta_1 \quad (4-8)$$

donde $\Theta_2 = \{\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \boldsymbol{\psi}^T, \boldsymbol{\lambda}^T)^T \mid \boldsymbol{\theta} \in \Theta_1, h_\tau(\boldsymbol{\lambda}) = 0, \tau = 1, \dots, \kappa\}$.

Los siguientes supuestos son establecidos:

(A1') Θ_1 es un conjunto compacto,

(A2') $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta})$ es una función continua y dos veces continuamente diferenciable en Θ_1

(A3') $\boldsymbol{\Sigma}(\mathbf{x}, \boldsymbol{\theta})$ es una función continua y dos veces continuamente diferenciable en Θ_1 .

Sean $\mathbf{y}_{k1}, \dots, \mathbf{y}_{kN_k}$ vectores de observaciones independientes con vector de tiempos de medición \mathbf{x}_k , $k = 1, \dots, s$. Bajo el modelo marginal (4-1), $\mathbf{y}_{k1}, \dots, \mathbf{y}_{kN_k}$, constituye una muestra aleatoria de vectores que provienen de una distribución $\mathcal{N}_n(\mathbf{f}(\mathbf{x}_k, \boldsymbol{\beta}), \boldsymbol{\Sigma}(\mathbf{x}_k, \boldsymbol{\theta}))$.

Para el problema de contraste (4-8), el logaritmo de la función de verosimilitud, con base en s muestras independientes, está dada por

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & \sum_{k=1}^s N_k \left[-\frac{n}{2} \log 2\pi - \frac{1}{2} \log \det(\boldsymbol{\Sigma}(\mathbf{x}_k, \boldsymbol{\theta})) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{x}_k, \boldsymbol{\theta}) \mathbf{S}_k) \right. \\ & \left. - \frac{1}{2} (\bar{\mathbf{y}}_k - \mathbf{f}(\mathbf{x}_k, \boldsymbol{\beta}))^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k, \boldsymbol{\theta}) (\bar{\mathbf{y}}_k - \mathbf{f}(\mathbf{x}_k, \boldsymbol{\beta})) \right], \end{aligned}$$

donde $\bar{\mathbf{y}}_k = N_k^{-1} \sum_{m=1}^{N_k} \mathbf{y}_{km}$ y $\mathbf{S}_k = N_k^{-1} \sum_{m=1}^{N_k} (\mathbf{y}_{km} - \bar{\mathbf{y}}_k)(\mathbf{y}_{km} - \bar{\mathbf{y}}_k)^T$. Si $\hat{\boldsymbol{\theta}}_1$ y $\hat{\boldsymbol{\theta}}_2$ son los estimadores de máxima verosimilitud en Θ_1 y Θ_2 respectivamente, entonces el estadístico de la razón de verosimilitud para (4-8), está dado por

$$\begin{aligned} -2 \log \hat{\lambda}_N &= -2 \{ \ell(\hat{\boldsymbol{\theta}}_2) - \ell(\hat{\boldsymbol{\theta}}_1) \} \\ &= \sum_{k=1}^s N_k \left\{ F_{ML}(\bar{\mathbf{y}}_k, \mathbf{S}_k, \hat{\mathbf{f}}_{2k}, \hat{\boldsymbol{\Sigma}}_{2k}) - F_{ML}(\bar{\mathbf{y}}_k, \mathbf{S}_k, \hat{\mathbf{f}}_{1k}, \hat{\boldsymbol{\Sigma}}_{1k}) \right\}, \end{aligned}$$

con

$$\begin{aligned} F_{ML}(\bar{\mathbf{y}}_k, \mathbf{S}_k, \hat{\mathbf{f}}_{jk}, \hat{\boldsymbol{\Sigma}}_{jk}) &= \log \det(\hat{\boldsymbol{\Sigma}}_{jk}) - \log \det(\mathbf{S}_k) - n + \text{tr} \left[\hat{\boldsymbol{\Sigma}}_{jk}^{-1} \mathbf{S}_k \right] \\ &+ (\bar{\mathbf{y}}_k - \hat{\mathbf{f}}_{jk})^T \hat{\boldsymbol{\Sigma}}_{jk}^{-1} (\bar{\mathbf{y}}_k - \hat{\mathbf{f}}_{jk}) \end{aligned} \quad (4-9)$$

donde $\hat{\mathbf{f}}_{jk} = \mathbf{f}(\mathbf{x}_k, \hat{\boldsymbol{\beta}}_j)$ y $\hat{\boldsymbol{\Sigma}}_{jk} = \boldsymbol{\Sigma}(\mathbf{x}_k, \hat{\boldsymbol{\theta}}_{jk})$, $j = 1, 2$.

Bajo H_0 , el estadístico de prueba $-2 \log \hat{\lambda}_N$ sigue asintóticamente una distribución χ^2 con κ grados de libertad. Por tanto, un test aproximado basado en $-2 \log \hat{\lambda}_N$ de tamaño α rechaza

H_0 si $-2 \log \widehat{\lambda}_N > c_{\kappa}(\alpha)$, donde $c_{\kappa}(\alpha)$ denota el punto porcentual $100\alpha\%$ superior de la distribución χ_{κ}^2 .

La función $F_{ML}(\bar{\mathbf{y}}_k, \mathbf{S}_k, \mathbf{f}_{jk}, \boldsymbol{\Sigma}_{jk})$ es conocida como la función de discrepancia de máxima verosimilitud la cual cuantifica la discrepancia entre los momentos muestrales y los momentos poblacionales que dependen del vector de parámetros $\boldsymbol{\theta}$ (ver Steiger *et al.*, 1985). La extensión de esta función, en el caso de s grupos, está definida por (ver Bollen, 1989):

$$F(\boldsymbol{\theta}) = \sum_{k=1}^s \frac{N_k}{N} F_{ML}(\bar{\mathbf{y}}_k, \mathbf{S}_k, \mathbf{f}_{jk}, \boldsymbol{\Sigma}_{jk}), \quad j = 1, 2. \quad (4-10)$$

Así, el estadístico (4-9) se puede escribir como

$$-2 \log \widehat{\lambda}_N = N(F(\widehat{\boldsymbol{\theta}}_2) - F(\widehat{\boldsymbol{\theta}}_1)). \quad (4-11)$$

Bajo los supuestos (A1')-(A3') y condiciones de regularidad se puede demostrar que la función criterio $T_W^{II}(\zeta_N)$ es proporcional al parámetro de no centralidad de la distribución χ_{κ}^2 del estadístico $-2 \log \widehat{\lambda}_N$ bajo una sucesión de hipótesis alternativas locales. Así la potencia de la prueba es una función monótona creciente de $T_W^{II}(\zeta_N)$ y por tanto se maximiza mediante la elección del diseño ζ_N .

A continuación se define el criterio T_W^{II} para diseños poblacionales aproximados.

Definición 4.1.2. *Asumiendo que \mathcal{M}_1 es el modelo correcto con vector de parámetros $\boldsymbol{\theta}_1^0$ conocido, un diseño poblacional aproximado T_W^{II} -óptimo local $\boldsymbol{\omega}^*$ para discriminar entre \mathcal{M}_1 y \mathcal{M}_2 verifica*

$$\boldsymbol{\omega}^* = \arg \max_{\boldsymbol{\omega} \in \Xi} T_W^{II}(\boldsymbol{\omega}) \quad (4-12)$$

donde

$$T_W^{II}(\boldsymbol{\omega}) = \min_{\boldsymbol{\theta}_2 \in \Theta_2} \sum_{i=1}^r \omega_i F(\mathbf{x}_i, \boldsymbol{\theta}_2) \quad (4-13)$$

con $F(\mathbf{x}_i, \boldsymbol{\theta}_2)$ dada por (4-7).

4.2. Teorema de equivalencia

Los resultados obtenidos en la sección 3.5 se pueden extender para el criterio T_W^{II} -optimal y se deducen de forma similar. Para establecer el teorema de equivalencia se adopta la siguiente definición de diseño T_W^{II} -regular.

Definición 4.2.1. *Un diseño $\boldsymbol{\omega}$ es llamado T_W^{II} -regular si el conjunto*

$$\Theta_2(\boldsymbol{\omega}) = \left\{ \tilde{\boldsymbol{\theta}}_2 : \tilde{\boldsymbol{\theta}}_2(\boldsymbol{\omega}) = \arg \min_{\boldsymbol{\theta}_2 \in \Theta_2} \sum_{i=1}^r \omega_i F(\mathbf{x}_i, \boldsymbol{\theta}_2) \right\}$$

es unitario, de lo contrario es llamado T_W^{II} -singular.

A continuación se enuncia el teorema de equivalencia para el criterio T_W^{II} -optimal cuya demostración es similar a la demostración del teorema 3.5.3.

Teorema 4.2.1. (Teorema de equivalencia para T_W^{II} -optimalidad). Sea ω^* un diseño T_W^{II} -regular. Entonces:

(i) Una condición necesaria y suficiente para que el diseño ω^* sea T_W^{II} -óptimo está dada por

$$\phi(\mathbf{x}, \omega^*) \leq 0, \quad \forall \mathbf{x} \in \bar{\mathcal{X}}^n, \quad \theta_2^* \in \Theta_2(\omega^*) \quad (4-14)$$

donde

$$\phi(\mathbf{x}, \omega) = F(\mathbf{x}, \theta_2^*) - T_W^{II}(\omega) \quad (4-15)$$

es la derivada direccional en ω^* en la dirección de δ_{ω_x} .

(ii) La función $\phi(\mathbf{x}, \omega^*)$ alcanza su valor máximo en los puntos de soporte del diseño ω^* .

(iii) El conjunto de diseños T_W^{II} -óptimos es convexo.

4.3. Discriminación en modelos no lineales de efectos mixtos

En esta sección se considera el problema de discriminación previamente descrito para el modelo general de efectos mixtos el cual se puede definir mediante un modelo jerárquico en dos etapas como se muestra a continuación.

4.3.1. Modelo no lineal de efectos mixtos

Suponga que el vector \mathbf{y}_i de respuestas para el i -ésimo individuo, $i = 1, \dots, N$ se representa por medio de un modelo no lineal de efectos mixtos, definido en dos etapas (ver Davidian y Giltinan, 1995):

Etapa 1. Modelo intra-individuo

En esta etapa se establece un modelo para la variabilidad entre las observaciones *dentro* de un individuo dado. Específicamente, para cada individuo i se asume que

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}_i) + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N, \quad (4-16)$$

donde $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^T$ es el vector que contiene los diferentes tiempos en los que se realizan las observaciones, $\boldsymbol{\beta}_i$ es un p -vector de parámetros asociados con el i -ésimo individuo, $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}_i) = (f(x_{i1}, \boldsymbol{\beta}_i), \dots, f(x_{in}, \boldsymbol{\beta}_i))^T$ y $\boldsymbol{\varepsilon}_i$ es el término del error aleatorio. En contraste

con los modelos marginales mixtos, la función $f(x_{il}, \boldsymbol{\beta}_i)$ es ahora una función no lineal del vector aleatorio $\boldsymbol{\beta}_i$.

Se asume que

$$\boldsymbol{\varepsilon}_i | \boldsymbol{\beta}_i \sim \mathcal{N}_n(\mathbf{0}, \text{Cov}(\boldsymbol{\varepsilon}_i | \boldsymbol{\beta}_i)), \quad \text{Cov}(\boldsymbol{\varepsilon}_i | \boldsymbol{\beta}_i) = \mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}_i, \sigma^2, \boldsymbol{\lambda}). \quad (4-17)$$

Así como en el caso de los modelos (3-1) y (4-1) la matriz \mathbf{R} representa la variabilidad intra-sujeto y se puede escribir en términos del modelo de varianza y la estructura de correlación como

$$\mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}_i, \sigma^2, \boldsymbol{\lambda}) = \mathbf{G}^{1/2}(\mathbf{x}_i, \boldsymbol{\beta}_i, \sigma^2, \boldsymbol{\alpha}) \mathbf{W}(\mathbf{x}_i, \boldsymbol{\rho}) \mathbf{G}^{1/2}(\mathbf{x}_i, \boldsymbol{\beta}_i, \sigma^2, \boldsymbol{\alpha}), \quad (4-18)$$

donde $\boldsymbol{\lambda} = (\boldsymbol{\alpha}^T, \boldsymbol{\rho}^T)^T$.

Dependiendo de la naturaleza de los datos, diferentes estructuras, incluyendo las descritas en la sección 3.1 y 4.1, se pueden utilizar para la matriz $\mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}_i, \sigma^2, \boldsymbol{\lambda})$.

Etapa 2. Modelo entre-individuos

Para describir la variabilidad *entre* individuos, en esta etapa se establece un modelo para el vector de parámetros $\boldsymbol{\beta}_i$ en función de ciertas características individuales. Como un caso especial, se supone que

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{b}_i, \quad (4-19)$$

donde $\boldsymbol{\beta}$ es un vector de parámetros poblacionales y \mathbf{b}_i es un vector de efectos aleatorios asociados con el i -ésimo individuo. Se asume que los vectores \mathbf{b}_i son independientes entre sí y de los vectores $\boldsymbol{\varepsilon}_i$, con distribución $\mathcal{N}_p(\mathbf{0}, \mathbf{D})$, donde \mathbf{D} es una matriz común para todos los individuos y es llamada la matriz de covarianzas entre-sujetos.

4.3.2. Modelo aproximado lineal

A partir de (4-16) y (4-19), y aproximando $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}_i)$ por una expansión de series de Taylor de primer orden alrededor del punto $E(\mathbf{b}_i) = \mathbf{0}$ se obtiene el modelo linealizado conocido como el *pseudo modelo no lineal de efectos mixtos de Beal y Sheiner*

$$\mathbf{y}_i \approx \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{Z}(\mathbf{x}_i, \boldsymbol{\beta}) \mathbf{b}_i + \mathbf{R}^{1/2}(\mathbf{x}_i, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}) \boldsymbol{\varepsilon}_i^*, \quad (4-20)$$

donde $\mathbf{Z}(\mathbf{x}_i, \boldsymbol{\beta}) = \left[\frac{\partial \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}_i)^T}{\partial \boldsymbol{\beta}_i} \Big|_{\boldsymbol{\beta}_i = \boldsymbol{\beta}} \right]$ es una matriz de orden $n \times p$ y $\boldsymbol{\varepsilon}_i^* \sim \text{i.i.d. } \mathcal{N}_n(0, \sigma^2 \mathbf{I}_n)$.

Se observa que este modelo aproximado corresponde a un modelo marginal mixto tipo II, no lineal en el vector $\boldsymbol{\beta}$ y lineal en los efectos aleatorios.

Además, se tiene que la distribución marginal de \mathbf{y}_i es aproximadamente normal con vector de medias

$$E(\mathbf{y}_i) \approx \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) \quad (4-21)$$

y matriz de varianzas y covarianzas

$$\begin{aligned}\text{Cov}(\mathbf{y}_i) &\approx \mathbf{Z}(\mathbf{x}_i, \boldsymbol{\beta}) \mathbf{D} \mathbf{Z}(\mathbf{x}_i, \boldsymbol{\beta})^T + \mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}) \\ &\equiv \boldsymbol{\Sigma}(\mathbf{x}_i, \boldsymbol{\theta}),\end{aligned}$$

que depende de $\boldsymbol{\beta}$. Así, asumiendo que la aproximación en (4-20) es exacta, la discriminación entre dos modelos de efectos mixtos anidados con respecto a la matriz de covarianzas intra-individuos se puede realizar con base en el diseño T_W^{II} -óptimo que se obtiene para discriminar entre los correspondientes modelos linealizados.

En la siguiente sección se muestra un ejemplo de aplicación del criterio para esta clase de modelos.

4.4. Un Ejemplo

Con el fin de ilustrar la aplicación del criterio en el caso de un modelo general de efectos mixtos se utilizó la información de un estudio de simulación sobre la farmacocinética de un medicamento realizado por Hashimoto *et al.* (1991). Este ejemplo también fue considerado por Mentré *et al.* (1995) quienes encontraron diseños poblacionales óptimos bajo dos modelos de variabilidad intra-sujeto con el propósito de estimar las características farmacocinéticas del medicamento. La siguiente es la información tomada del segundo estudio.

Modelos y condiciones experimentales

Para describir la evolución de las concentraciones del medicamento en el organismo en función del tiempo se asumió el modelo abierto de un compartimiento con dosis única \mathcal{D} dado por

$$\text{concentración en plasma} = \frac{\mathcal{D}}{V} \exp\left(-\frac{Cl}{V} \times \text{tiempo}\right)$$

donde V y Cl son los parámetros del modelo y representan en la cinética de un fármaco el volumen de distribución y el aclaramiento, respectivamente. Debido a que el objetivo era estudiar las características poblacionales del medicamento (parámetros medios y variabilidad entre-individuos) con un número limitado de muestras por sujeto se consideró el siguiente modelo no lineal de efectos mixtos:

Etapa 1. (Variación intra-individuo)

$$\begin{aligned}y_{il} &= f(x_{il}, \boldsymbol{\beta}_i) + \varepsilon_{il} \\ f(x_{il}, \boldsymbol{\beta}_i) &= \frac{\mathcal{D}}{V_i} \exp\left(-\frac{Cl_i}{V_i} x_{il}\right) \\ \varepsilon_i | \boldsymbol{\beta}_i &\sim N_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}_i, \sigma^2, \boldsymbol{\lambda}))\end{aligned}\tag{4-22}$$

donde, para el sujeto i , y_{il} representa la l -ésima concentración en el tiempo x_{il} y $\beta_i = (Cl_i, V_i)^T$. Para todos los individuos se asumió una dosis igual a 1.

Etapa 2. (Variación entre-individuos)

$$\beta_i = \beta + \mathbf{b}_i, \quad \mathbf{b}_i \sim N_2(\mathbf{0}, \mathbf{D}), \quad \mathbf{D} = \begin{bmatrix} \psi_{\beta_1} & 0 \\ 0 & \psi_{\beta_2} \end{bmatrix}, \quad (4-23)$$

donde $\beta = (\beta_1, \beta_2)^T$ es el vector de parámetros poblacionales.

Los dos modelos considerados para la matriz de covarianzas intra-individuos $R(\mathbf{x}_i, \beta_i, \sigma^2, \lambda)$ fueron:

\mathcal{M}_1 . Errores no correlacionados con varianza proporcional a una potencia de la respuesta media

$$\begin{aligned} \varepsilon_i | \beta_i &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \beta_i, \sigma^2, \lambda)) \\ \mathbf{R}(\mathbf{x}_i, \beta_i, \sigma^2, \lambda) &= \sigma^2 \mathbf{G}(\mathbf{x}_i, \beta_i, \alpha), \end{aligned}$$

donde

$$\mathbf{G}(\mathbf{x}_i, \beta_i, \alpha) = \text{diag}(f^{2\alpha}(x_{i1}, \beta_i), \dots, f^{2\alpha}(x_{in}, \beta_i))$$

\mathcal{M}_2 . Errores no correlacionados con varianza constante

$$\begin{aligned} \varepsilon_i | \beta_i &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \beta_i, \sigma^2, \lambda)) \\ \mathbf{R}(\mathbf{x}_i, \beta_i, \sigma^2, \lambda) &= \sigma^2 \mathbf{I}_n. \end{aligned}$$

Para la construcción de los diseños D-óptimos locales se tomó

$$\mathcal{X} = \{0.05, 0.15, 0.3, 0.6, 1\} \text{ horas} \quad (4-24)$$

y se usaron como vectores de parámetros locales

$$\beta = (0.5, 0.2)^T, \quad \psi_1 = (0.01, 0.0016)^T \text{ y } \alpha = 1. \quad (4-25)$$

En ambos modelos se asumió $\sigma_1^2 = \sigma_2^2 = 0.15$.

Diseño T_W^{II} -óptimo

A partir de la información anterior, en esta sección, se construye el diseño T_W^{II} -óptimo local con el fin de discriminar entre \mathcal{M}_1 y \mathcal{M}_2 . Debido a que \mathcal{M}_2 está anidado dentro de \mathcal{M}_1 , se asume que \mathcal{M}_1 es el modelo correcto con parámetros dados por (4-25).

Para el modelo alternativo \mathcal{M}_2 el vector de parámetros poblacionales es

$$\theta_2 = (\beta_2^T, \sigma_2^2, \psi_2^T, 0)^T. \quad (4-26)$$

Con fines ilustrativos se fijaron tres muestras por individuo, es decir, $n = 3$. De acuerdo con los resultados obtenido por Mentré *et al.* (1995), este número en experimentos con muestras escasas es razonable. Así, la región de diseño $\bar{\mathcal{X}}^3$ contiene $C_{5,3} = 10$ combinaciones posibles de secuencias de elementos tomados de \mathcal{X} :

$$\bar{\mathcal{X}}^3 = \{(x_1, x_2, x_3) : x_s \in \mathcal{X}\}. \quad (4-27)$$

El diseño poblacional óptimo local resultante es el diseño de dos puntos dado por

$$\zeta_{TW}^* = \left\{ \begin{array}{cc} (0.05, 0.15, 0.3) & (0.05, 0.15, 0.6) \\ 0.82 & 0.18 \end{array} \right\} \quad (4-28)$$

y el vector de parámetros del modelo \mathcal{M}_2 obtenido en el proceso de optimización es

$$\begin{aligned} \theta_2^* &= (Cl_2^*, V_2^*, \psi_{Cl_1}^*, \psi_{V_1}^*)^T \\ &= (0.6287, 0.1998, 1.4059, 0.0027)^T. \end{aligned}$$

Así, si el objetivo es diseñar un experimento con muestras escasas por individuo con el fin de discriminar de forma óptima entre los modelos \mathcal{M}_1 y \mathcal{M}_2 entonces las muestras de sangre se deben tomar en los tiempos 0.05, 0.15 y 0.3 (horas) para aproximadamente el 82 % de los pacientes y en los tiempos 0.05, 0.15 and 0.6 (horas) para aproximadamente el 18 % de los pacientes. En la gráfica 4-1 se verifica que en efecto el diseño es T_W^{II} -óptimo.

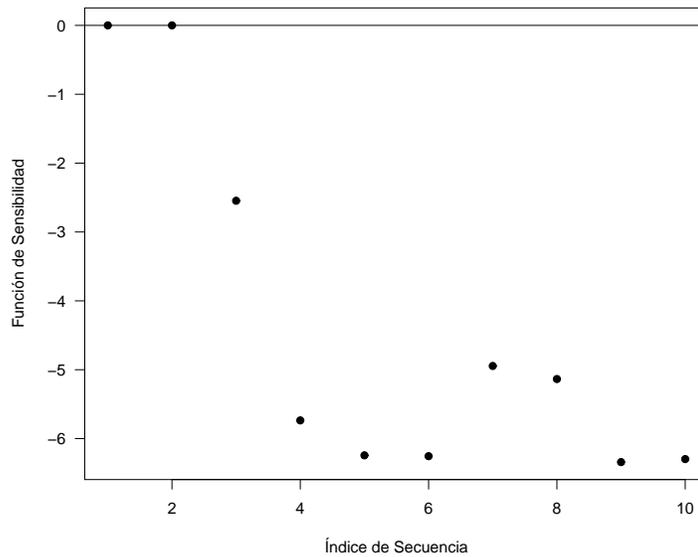


Figura 4-1.: Función de sensibilidad para los vectores de muestreo.

Comparación con otros diseños

En la Tabla 4-1 se muestran los diseños D-óptimos para cada modelo, el diseño producto y un diseño de rejilla de segundo orden con secuencias de medición (x_1, x_3, x_5) y (x_2, x_4) y la misma proporción de individuos en cada una de ellas.

Tabla 4-1.: T_W^{II} -eficiencias de diseños poblacionales

Diseño	$T_W^{II} - \text{ef}(\zeta)$
$\zeta_{T_W^{II}}^* = \left\{ \begin{array}{cc} (0.05, 0.15, 0.30) & (0.05, 0.15, 0.60) \\ 0.82 & 0.18 \end{array} \right\}$	1
$\zeta_P^* = \left\{ \begin{array}{cc} (0.05, 0.30, 0.60) & (0.05, 0.30, 1) \\ 0.01 & 0.99 \end{array} \right\}$	0.2609
$\zeta_{D_1}^* = \left\{ \begin{array}{c} (0.05, 0.15, 1) \\ 1 \end{array} \right\}$	0.4601
$\zeta_{D_2}^* = \left\{ \begin{array}{c} (0.05, 0.30, 0.60) \\ 1 \end{array} \right\}$	0.2876
$\zeta_R = \left\{ \begin{array}{cc} (0.05, 0.30, 1) & (0.15, 0.60) \\ 0.50 & 0.50 \end{array} \right\}$	0.2604

Las T_W^{II} -eficiencias de estos diseños se encuentran en la Tabla 4-1. Se observa que la T_W^{II} -eficiencia máxima es tan sólo del 46% correspondiente al diseño $\zeta_{D_1}^*$. Los demás diseños, incluyendo el diseño producto, tienen eficiencias inferiores al 29% con respecto al diseño $\zeta_{T_W^{II}}^*$ -óptimo.

5. Diseños óptimos para estimación y discriminación en modelos mixtos

En este capítulo se propone una metodología para construir diseños óptimos poblacionales con dos objetivos simultáneos: estimación de parámetros y discriminación entre dos estructuras de variabilidad intra-sujetos para modelos de efectos mixtos. Se propone un criterio compuesto que combina un criterio para estimación y otro para discriminación. Para estimación se usa el criterio de D-optimalidad, ampliamente utilizado en el caso de modelos de efectos mixtos (ver por ejemplo Mentré *et al.*, 1997; Gagnon y Leonov, 2005; entre otros) y para discriminación entre modelos se considera el criterio T_W propuesto en este trabajo. El principal resultado de este capítulo es la formulación del teorema de equivalencia para el criterio compuesto mediante la aplicación de la teoría de diseños aproximados.

El capítulo se desarrolla en cinco secciones. En la sección 5.1 se presenta la descripción del problema de estimación y discriminación. En la sección 5.2 se presenta una expresión para la matriz de información Fisher para modelos no lineales de efectos mixtos. En la sección 5.3 se describen los diseños D -óptimos poblacionales y se enuncia el respectivo teorema de equivalencia. En la sección 5.4 se construye un criterio de optimalidad para la estimación de parámetros y discriminación entre dos modelos de efectos mixtos. Se enuncia y se demuestra el respectivo teorema de equivalencia. Finalmente, en la sección 5.5 se muestra un ejemplo de aplicación.

5.1. Descripción del problema de estimación y discriminación

Diferentes áreas de aplicación de los modelos mixtos, principalmente la farmacocinética, han impulsado el desarrollo de varios enfoques para la construcción de diseños poblacionales óptimos, especialmente con el propósito de obtener estimadores eficientes de los parámetros del modelo. En contraste con la estimación de parámetros, no hay muchos métodos que aborden el problema de discriminación entre modelos. En el caso de modelos de efectos fijos, algunos autores han puntualizado que los diseños óptimos obtenidos para discriminación, en general, no son eficientes para la estimación de los parámetros del modelo (ver por ejemplo Atkinson y Donev, 1992). Lo ideal sería construir un diseño que sea eficiente tanto para estimación como para discriminación. En el contexto de modelos de efectos fijos, varios

enfoques se han desarrollado con el fin de construir diseños con ambos propósitos. Algunos resultados de criterios compuestos aparecen en Atkinson (2008), quien propuso el criterio DT-optimalidad, López-Río (2008) consideró el problema de diseños óptimos para la estimación de funciones no lineales y discriminación, Waterhouse *et al.* (2005) propusieron un criterio producto para discriminación y estimación basado en el producto de los determinantes de las matrices de información. Otros enfoques consideran la construcción de diseños híbridos y condicionales. Para modelos no lineales de efectos mixtos el problema de diseños para estimación y discriminación ha sido considerado por algunos autores. Por ejemplo, Jones y Wang (1999) sugirieron el uso de diseños compuestos combinando T-optimalidad con el criterio D-optimalidad para modelos de efectos mixtos, Waterhouse *et al.* (2005) sugirieron el uso del criterio producto y Vajjah y Duffull (2012) propusieron diseños condicionales combinando D-optimalidad para modelos mixtos y T-optimalidad para las componentes fijas.

En este trabajo el objetivo es proponer un criterio de optimalidad con el fin de obtener diseños óptimos que permitan discriminar entre dos estructuras de covarianza intra-individuo y además estimar de forma eficiente los parámetros de cada modelo. El criterio propuesto es un criterio de diseño compuesto que combina D-optimalidad para modelos mixtos y T_W -optimalidad para discriminación según el modelo marginal considerado. Aunque, enfoques como los descritos anteriormente resuelven el problema de diseño para estimación de parámetros y discriminación entre modelos más generales, la metodología que se desarrolla en este trabajo permite la discriminación entre modelos anidados, la posibilidad de grupos de individuos con diferentes diseños individuales y la verificación de la optimalidad de un diseño.

Para describir el problema de diseños óptimos para estimación de parámetros en modelos no lineales de efectos mixtos, se presenta a continuación una expresión para la matriz de información de Fisher.

5.2. Matriz de Información de Fisher

Considere el modelo marginal (4-1) con vector de parámetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T \in \Theta \subset \mathbb{R}^v$, el cual se puede expresar de forma compacta como

$$\mathbf{y}_i \sim \mathcal{N}_n(\mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta}), \boldsymbol{\Sigma}(\mathbf{x}_i, \boldsymbol{\theta})), \quad i = 1, \dots, N, \quad (5-1)$$

donde la función $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta})$ depende solamente del vector de parámetros $\boldsymbol{\beta}$.

Bajo ciertas condiciones de regularidad, el estimador de máxima verosimilitud de $\boldsymbol{\theta}$ con base en la muestra total de N individuos tiene una distribución asintótica normal

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}_v(\boldsymbol{\theta}^0, N^{-1}M^{-1}(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\theta}^0)) \quad (5-2)$$

donde $M(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\theta})$ es la matriz de información de Fisher dada por

$$M(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\theta}) = \sum_{i=1}^N \boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta}) \quad (5-3)$$

con

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{x}_i, \boldsymbol{\theta}) &= \{\boldsymbol{\mu}_{kl}(\mathbf{x}_i, \boldsymbol{\theta})\}_{k,l=1}^v \\ \boldsymbol{\mu}_{kl}(\mathbf{x}_i, \boldsymbol{\theta}) &= \frac{\partial \mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}_k} \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i, \boldsymbol{\theta}) \frac{\partial \mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_l} \\ &\quad + \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i, \boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i, \boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_l} \right], \end{aligned} \quad (5-4)$$

ver Magnus y Neudecker (1988).

Las condiciones de regularidad requeridas son (ver Cox y Hinkley, 1974, Capítulo 9, p. 281):

- (i) Θ es un conjunto compacto.
- (ii) $\boldsymbol{\theta}^0$ es un punto interior de Θ .
- (iii) Valores diferentes del vector de parámetros $\boldsymbol{\theta}$ corresponden a distribuciones diferentes.
- (iv) La función log-verosimilitud es tres veces diferenciable con respecto a $\boldsymbol{\theta}$ en una vecindad de $\boldsymbol{\theta}^0$.
- (v) El valor absoluto de la tercera derivada de la función log-verosimilitud con respecto a $\boldsymbol{\theta}$ está acotada por una función de \mathbf{y} con valor esperado finito, en una vecindad de $\boldsymbol{\theta}^0$.
- (vi) La matriz de información de Fisher es finita y definida positiva.

En el caso general de los modelos no lineales de efectos mixtos, no es posible obtener una expresión cerrada para la matriz de información de Fisher, debido a que la función de respuesta es no lineal en los efectos aleatorios. Un enfoque comúnmente usado para obtener expresiones aproximadas del vector de medias y la matriz de covarianzas de la distribución marginal, es considerar el modelo aproximado lineal como se describió en la sección 4.3.2. Así, bajo este enfoque se tiene que:

$$\begin{aligned} E(\mathbf{y}_i) &\approx \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) \\ \text{Cov}(\mathbf{y}_i) &\approx \mathbf{Z}(\mathbf{x}_i, \boldsymbol{\beta}) \mathbf{D} \mathbf{Z}(\mathbf{x}_i, \boldsymbol{\beta})^T + \mathbf{R}(\mathbf{x}_i, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}) \\ &\equiv \boldsymbol{\Sigma}(\mathbf{x}_i, \boldsymbol{\theta}), \end{aligned}$$

y una expresión aproximada para la matriz de información de Fisher se obtiene a partir de (5-4).

De la expresión en (5-2), se tiene que asintóticamente la matriz de covarianzas del estimador de máxima verosimilitud de $\boldsymbol{\theta}$ alcanza la cota inferior de la desigualdad de Cramér-Rao. Esta

propiedad es lo que motiva el uso de la inversa de la matriz (5-3) como una aproximación de la matriz de covarianzas del estimador y en consecuencia emplearla como una herramienta para evaluar la calidad o precisión del estimador. Entre más “pequeña” sea esta matriz más precisa será la estimación de $\boldsymbol{\theta}$.

A partir de la expresión (5-3) se observa que la matriz de información depende de las condiciones experimentales \boldsymbol{x} por tanto el problema de diseño consiste en elegir el diseño poblacional exacto de tamaño N que hace “grande” esta matriz en algún sentido.

Cada diseño poblacional genera una matriz de información de Fisher la cual puede escribirse como

$$M_N(\boldsymbol{\theta}) = \sum_{k=1}^s N_k \boldsymbol{\mu}(\boldsymbol{x}_k, \boldsymbol{\theta}) = NM(\zeta_N, \boldsymbol{\theta}) \quad (5-5)$$

donde

$$M(\zeta_N, \boldsymbol{\theta}) = \sum_{k=1}^s \frac{N_k}{N} \boldsymbol{\mu}(\boldsymbol{x}_k, \boldsymbol{\theta}) = \sum_{k=1}^s \omega_k \boldsymbol{\mu}(\boldsymbol{x}_k, \boldsymbol{\theta}). \quad (5-6)$$

Así, para N individuos el objetivo es elegir el diseño ζ_N que minimice $Cov(\hat{\boldsymbol{\theta}}) \propto M^{-1}(\zeta_N, \boldsymbol{\theta})$ en algún sentido. Como se mencionó en la sección 2.3, el enfoque utilizado para minimizar esta matriz es maximizar un funcional de la matriz de información de Fisher cuya elección dependerá de los intereses del experimento. Es decir, se busca el diseño tal que

$$\zeta_N^* = \arg \max_{\zeta_N} \Phi(M(\zeta_N, \boldsymbol{\theta})), \quad (5-7)$$

donde $\Phi(M(\zeta_N, \boldsymbol{\theta}))$ es una función acorde con las necesidades del experimento. Con el fin de simplificar este problema y proporcionar condiciones para verificar la Φ -optimalidad de un diseño se usan los diseños poblacionales aproximados continuos.

Generalizando la expresión (5-6), se define la matriz de información de Fisher para un diseño poblacional continuo $\boldsymbol{\omega}$ como

$$M(\boldsymbol{\omega}, \boldsymbol{\theta}) = \sum_{k=1}^r \omega_k \boldsymbol{\mu}(\boldsymbol{x}_k, \boldsymbol{\theta}). \quad (5-8)$$

El conjunto de las matrices de información asociadas a los diseños poblacionales continuos en Ξ se denota por \mathcal{M} y está dado por

$$\mathcal{M} = \{M(\boldsymbol{\omega}, \boldsymbol{\theta}) : \boldsymbol{\omega} \in \Xi\}. \quad (5-9)$$

A continuación se presentan algunas propiedades de la matriz de información. El teorema se demuestra de forma similar al teorema establecido en Fedorov (1972), Capítulo 5, p.210.

Teorema 5.2.1. *Para cualquier diseño $\omega \in \Xi$ se tiene que:*

- (i) *La matriz $M(\omega, \theta)$ es simétrica y definida no negativa.*
- (ii) *Bajo los supuestos (A1')-(A3') se verifica que el conjunto \mathcal{M} es compacto y convexo.*
- (iii) *Para cualquier matriz M de \mathcal{M} existe un diseño ω que no contiene más de $v(v+1)/2+1$ puntos con pesos distintos de cero, de tal manera que $M(\omega, \theta) = M$.*

En la parte (ii) del teorema es importante notar que la convexidad del conjunto se sigue del hecho de que éste coincide con la cubierta convexa del conjunto de las matrices de información de los diseños individuales en Ξ , es decir

$$\mathcal{M} = \text{conv}(\mathcal{M}_{\text{ind}}), \quad (5-10)$$

donde

$$\mathcal{M}_{\text{ind}} = \{\mu(\mathbf{x}, \theta) : \mathbf{x} \in \bar{\mathcal{X}}^n\}. \quad (5-11)$$

5.3. Diseños óptimos para estimación

Con el fin de encontrar diseños para la estimación eficiente del vector de parámetros θ , se considera el criterio de D-optimalidad definido en el contexto de los diseños poblacionales como sigue.

Definición 5.3.1. *Un diseño poblacional ω^* se dice que es D-óptimo local si verifica,*

$$\omega^* = \arg \max_{\omega \in \Xi} \Phi_D(M(\omega, \theta)) \quad (5-12)$$

donde $\Phi_D(M(\omega, \theta)) = \log \det M(\omega, \theta)$.

Las siguientes son algunas propiedades del criterio Φ_D :

- (i) La función toma un valor finito si y sólo si $M(\omega, \theta)$ es una matriz no singular.
- (ii) La función es continua en \mathcal{M} .
- (iii) Φ_D es una función cóncava en \mathcal{M} y estrictamente cóncava en el conjunto de las matrices definidas positivas.
- (iv) La función es diferenciable en el subconjunto de las matrices definidas positivas y su gradiente es

$$\nabla \Phi_D(M) = M^{-1}. \quad (5-13)$$

- (v) Bajo el supuesto de normalidad, los diseños que maximizan Φ_D minimizan el volumen del elipsoide de confianza para el vector de parámetros.

El siguiente es el teorema de equivalencia para D-optimalidad.

Teorema 5.3.1. (Teorema de equivalencia para D-optimalidad). *Sea $\omega^* \in \Xi$ un diseño poblacional con matriz de información de Fisher no singular $M(\omega^*, \theta)$. Entonces una condición necesaria y suficiente para que el diseño ω^* sea D-óptimo local está dada por:*

$$\phi_D(\mathbf{x}, \omega^*) \leq 0, \quad \forall \mathbf{x} \in \bar{\mathcal{X}}^n, \quad (5-14)$$

donde

$$\phi_D(\mathbf{x}, \omega^*) = \text{tr} [\boldsymbol{\mu}(\mathbf{x}, \theta) M^{-1}(\omega^*, \theta)] - v, \quad v = \dim(\theta) \quad (5-15)$$

es la derivada direccional de $\Phi_D(M(\omega, \theta))$ en ω en la dirección de $\omega_{\mathbf{x}}$. Además la función $\phi_D(\mathbf{x}, \omega^*)$ alcanza su valor máximo en los puntos de soporte del diseño ω^* .

Demostración. Sea $H(\omega, \theta)$ una matriz simétrica y $M(\omega, \theta)$ una matriz no singular. Por las propiedades anteriores Φ_D es una función cóncava en \mathcal{M} y diferenciable en el conjunto de matrices definidas positivas. Luego a partir de (2-16) y la propiedad (5-13) se tiene que la derivada direccional de Φ_D en M en la dirección de H está dada por

$$F_{\Phi_D}(M, H) = \text{tr}[\nabla \Phi_D(M)(H - M)] = \text{tr}(M^{-1}H) - v. \quad (5-16)$$

Así, aplicando el teorema 2.5.3 y el corolario 2.5.4.1 se tiene que ω^* es D-optimal si y sólo si M es no singular y $F_{\Phi_D}(M(\omega^*, \theta), \mu(\mathbf{x}, \theta) - M(\omega^*, \theta)) \leq 0$, lo cual es equivalente a:

$$\text{tr}(M^{-1}(\omega^*, \theta)\mu(\mathbf{x}, \theta)) - v \leq 0, \quad \forall \mathbf{x} \in \bar{\mathcal{X}}^n \quad (5-17)$$

con la igualdad en los puntos de soporte del diseño ω^* .

□

5.4. Diseños para estimación y discriminación

En esta sección se propone un criterio de optimalidad a partir de los criterios expuestos para discriminar entre dos modelos de covarianza intra-sujetos y estimar de forma eficiente el vector de parámetros.

Sean \mathcal{M}_1 y \mathcal{M}_2 dos modelos de efectos mixtos con vectores de parámetros θ_1 y θ_2 donde $\theta_1 \in \Theta_1 \subset \mathbb{R}^{v_1}$ y $\theta_2 \in \Theta_2 \subset \mathbb{R}^{v_2}$, los cuales asumen la misma estructura para la función de respuesta media $\mathbf{f}(\mathbf{x}, \beta)$ y matriz de covarianzas entre-individuos \mathbf{D} (correctamente especificadas), pero modelos diferentes para la variabilidad intra-individual. Además, suponga que

los modelos están anidados con respecto a la variabilidad intra-sujetos. En forma compacta los modelos se pueden representar como:

$$\mathcal{M}_1 : \varepsilon_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}_1(\mathbf{x}_i, \boldsymbol{\beta}_1, \sigma_1^2, \boldsymbol{\lambda}_1)), \quad \boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1^T, \boldsymbol{\gamma}_1^T)^T \in \Theta_1$$

$$\boldsymbol{\gamma}_1 = (\sigma_1^2, \boldsymbol{\psi}_1^T, \boldsymbol{\lambda}_1^T)^T \in \Gamma_1 \subset \mathbb{R}^{m_1}$$

$$\mathcal{M}_2 : \varepsilon_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}_2(\mathbf{x}_i, \boldsymbol{\beta}_2, \sigma_2^2, \boldsymbol{\lambda}_2)), \quad \boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2^T, \boldsymbol{\gamma}_2^T)^T \in \Theta_2$$

$$\boldsymbol{\gamma}_2 \in \Gamma_2 = \{\boldsymbol{\gamma}_2 = (\sigma_2^2, \boldsymbol{\psi}_2^T, \boldsymbol{\lambda}_2^T)^T \in \Gamma_1 : h_\tau(\boldsymbol{\lambda}_2) = 0, \tau = 1, \dots, \kappa\}.$$

El objetivo es proponer un criterio de optimalidad que permita encontrar diseños óptimos locales con el fin de:

1. Discriminar entre las estructuras de covarianza $\mathbf{R}_1(\mathbf{x}_i, \boldsymbol{\beta}_1, \sigma_1^2, \boldsymbol{\lambda}_1)$ y $\mathbf{R}_2(\mathbf{x}_i, \boldsymbol{\beta}_2, \sigma_2^2, \boldsymbol{\lambda}_2)$, y
2. Estimar en forma óptima $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}_2$.

Los criterios considerados para discriminación y estimación son los siguientes:

- $\Phi_1(\boldsymbol{\omega}) = T_W(\boldsymbol{\omega})$ para discriminación asumiendo que el modelo correcto es \mathcal{M}_1 .
- $\Phi_2(\boldsymbol{\omega}) = \det M_1(\boldsymbol{\omega})^{1/v_1}$ para estimación del vector de parámetros $\boldsymbol{\theta}_1$.
- $\Phi_3(\boldsymbol{\omega}) = \det M_2(\boldsymbol{\omega})^{1/v_2}$ para estimación del vector de parámetros $\boldsymbol{\theta}_2$.

donde $M(\boldsymbol{\omega})$ representa la matriz de información de Fisher para un valor a priori $\boldsymbol{\theta}_0$ de $\boldsymbol{\theta}$.

El criterio propuesto es un criterio compuesto que usa la media geométrica de las eficiencias de un diseño $\boldsymbol{\omega}$ con respecto a cada criterio como función apropiada para combinar los distintos criterios bajo ciertas ponderaciones. Como en López-Ríos (2008), en este trabajo se considera asignar la misma ponderación a los criterios para estimación y otra ponderación para discriminación.

Para una constante predefinida α donde $0 \leq \alpha \leq 1$, considere la función:

$$\Phi(\boldsymbol{\omega}) = \left\{ \frac{\Phi_1(\boldsymbol{\omega})}{\Phi_1(\boldsymbol{\omega}_1^*)} \right\}^\alpha \left\{ \frac{\Phi_2(\boldsymbol{\omega})}{\Phi_2(\boldsymbol{\omega}_2^*)} \frac{\Phi_3(\boldsymbol{\omega})}{\Phi_3(\boldsymbol{\omega}_3^*)} \right\}^{\frac{1-\alpha}{2}} \quad (5-18)$$

donde $\boldsymbol{\omega}_j^* = \arg \max_{\boldsymbol{\omega} \in \Xi} \Phi_j(\boldsymbol{\omega})$. El valor de la constante α se puede elegir de acuerdo con la importancia que se le quiera asignar a estimación y discriminación. También se pueden emplear métodos para determinar esta constante (ver por ejemplo Atkinson y Donev, 1992). El logaritmo de $\Phi(\boldsymbol{\omega})$ es

$$\Phi_\alpha^{DTW}(\boldsymbol{\omega}) = \alpha \ln T_W(\boldsymbol{\omega}) + \frac{1-\alpha}{2} \left\{ \frac{1}{v_1} \ln \det M_1(\boldsymbol{\omega}) + \frac{1}{v_2} \ln \det M_2(\boldsymbol{\omega}) \right\} + C \quad (5-19)$$

donde C es un constante que no depende del diseño ω . Así, se propone el siguiente criterio de optimalidad.

Definición 5.4.1. *Asumiendo que \mathcal{M}_1 es el modelo correcto, un diseño poblacional óptimo local ω^* para discriminar entre \mathcal{M}_1 y \mathcal{M}_2 , y estimar los vectores de parámetros θ_1 y θ_2 , es tal que*

$$\omega^* = \arg \max_{\omega \in \Xi} \Phi_\alpha^{DTW}(\omega) \quad (5-20)$$

donde $\Phi_\alpha^{DTW}(\omega)$ está dado por (5-19). El diseño ω^* es llamado DT_W^α -óptimo.

A continuación se enuncia y se demuestra el respectivo teorema de equivalencia.

Teorema 5.4.1. (Teorema de equivalencia para DT_W^α -óptimalidad). *Sea ω^* un diseño T_W -regular con matriz de información de Fisher no singular. Entonces bajo los supuestos (A1')-(A3') se tiene que una condición necesaria y suficiente para que el diseño ω^* sea DT_W^α -óptimo local está dada por*

$$\phi_\alpha^{DTW}(\mathbf{x}, \omega^*) \leq 0, \quad \forall \mathbf{x} \in \bar{\mathcal{X}}^n, \quad (5-21)$$

donde

$$\begin{aligned} \phi_\alpha^{DTW}(\mathbf{x}, \omega^*) &= \alpha \frac{F(\mathbf{x}, \theta_2^*)}{T_W^I(\omega^*)} + \frac{1-\alpha}{2} \left\{ \frac{1}{v_1} \text{tr} [\boldsymbol{\mu}_1(\mathbf{x}) M_1^{-1}(\omega^*)] \right. \\ &\quad \left. + \frac{1}{v_2} \text{tr} [\boldsymbol{\mu}_2(\mathbf{x}) M_2^{-1}(\omega^*)] \right\} - 1 \end{aligned} \quad (5-22)$$

es la derivada direccional de $\Phi_\alpha^{DTW}(\omega)$ en ω^* en la dirección de ω_x . Además, la función $\phi_\alpha^{DTW}(\mathbf{x}, \omega^*)$ alcanza su valor máximo en los puntos de soporte del diseño ω^* .

Demostración. Debido a que el criterio $\Phi_\alpha^{DTW}(\omega)$ es una combinación convexa de criterios cóncavos se tiene que éste es una función cóncava. Además es diferenciable ya que cada $\Phi_j(\omega)$ es diferenciable. A partir de las propiedades de la derivada direccional se sigue que la derivada direccional de $\Phi_\alpha^{DTW}(\omega)$ en ω^* en la dirección de ω_x es:

$$F_{\Phi_\alpha^{DTW}}(\omega^*, \omega_x) = \alpha F_{\Phi_1}(\omega^*, \omega_x) + \frac{1-\alpha}{2} [F_{\Phi_2}(\omega^*, \omega_x) + F_{\Phi_3}(\omega^*, \omega_x)] \quad (5-23)$$

donde

$$F_{\Phi_2}(\omega^*, \omega_x) = \frac{1}{v_1} \{ \text{tr} [\boldsymbol{\mu}_1(\mathbf{x}) M_1^{-1}(\omega^*)] - v_1 \}, \quad (5-24)$$

$$F_{\Phi_3}(\omega^*, \omega_x) = \frac{1}{v_2} \{ \text{tr} [\boldsymbol{\mu}_2(\mathbf{x}) M_2^{-1}(\omega^*)] - v_2 \} \quad (5-25)$$

y para T_W -óptimalidad, aplicando la regla de la cadena se obtiene que:

$$F_{\Phi_1}(\omega^*, \omega_x) = \frac{F(\mathbf{x}, \theta_2^*) - T_W(\omega^*)}{T_W(\omega^*)}. \quad (5-26)$$

Reemplazando estas expresiones y simplificando se obtiene la expresión del lado izquierdo de (5-22). El resultado se sigue aplicando el teorema 2.5.5. □

5.5. Un ejemplo

Con el fin de ilustrar la metodología propuesta, se considera un estudio realizado sobre la farmacocinética del medicamento cadralazina utilizado como tratamiento para la hipertensión, ver Racine *et al.* (1986). Los datos corresponden a las concentraciones de cadralazina en plasma en 10 pacientes con insuficiencia cardíaca. Los niveles plasmáticos fueron observados en un periodo de 2 a 32 horas después de una dosis única de 30 mg de cadralazina.

Estos datos se representan gráficamente en la Figura 5-1.

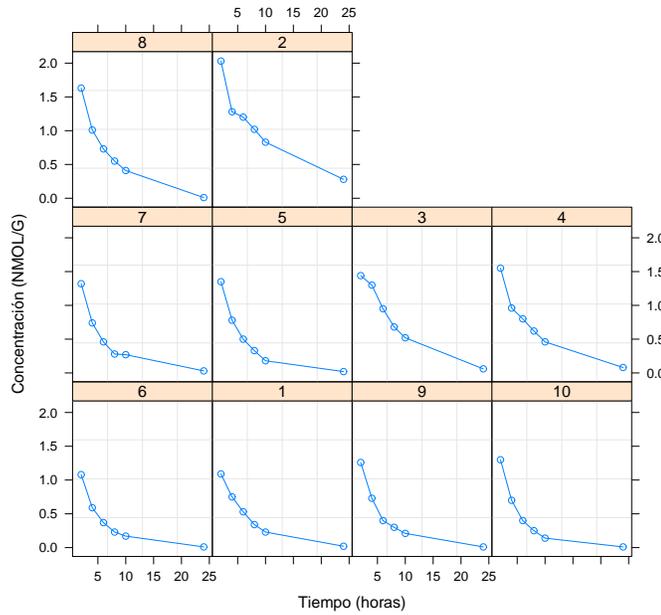


Figura 5-1.: Concentración de Cadralazina.

Para el estudio de la cadralazina, se considera el siguiente modelo de intercepto aleatorio:

$$y_{il} = \frac{D}{\beta_1} \exp(-\beta_2 x_{il}) + b_i + \varepsilon_{il}, \quad (5-27)$$

donde b_i es el efecto aleatorio asociado al i -ésimo individuo y ε_{il} es el término del error aleatorio intra-individuo. Sea $\mathbf{b} = (b_1, \dots, b_N)^T$ y $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in})^T$. Asuma que:

$$\mathbf{b} \sim \mathcal{N}_N(\mathbf{0}, \psi \mathbf{I}_N), \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})), \quad Cov(\mathbf{b}, \boldsymbol{\varepsilon}_i) = \mathbf{0}. \quad (5-28)$$

Las dos estructuras consideradas para la variabilidad intra-individual son: (i) un modelo de varianza con función exponencial y un modelo autorregresivo continuo con el tiempo como covariable y (ii) la estructura usual: errores independientes con varianza constante. En forma compacta, estos modelos se pueden escribir como:

\mathcal{M}_1 . *Función de varianza exponencial y modelo AR(1) continuo*

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})), \quad \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) = \mathbf{G}^{1/2}(\mathbf{x}_i, \sigma^2, \varpi) \mathbf{W}(\mathbf{x}_i, \rho) \mathbf{G}^{1/2}(\mathbf{x}_i, \sigma^2, \varpi)$$

donde

$$\begin{aligned} \mathbf{G}(\mathbf{x}_i, \sigma^2, \varpi) &= \sigma^2 \text{diag}(\exp(-\varpi x_{i1}), \dots, \exp(-\varpi x_{in})) \\ \mathbf{W}(\mathbf{x}_i, \rho) &= \{w_{sk}(\mathbf{x}_i, \rho)\}_{s,k=1}^n, \quad w_{sk}(\mathbf{x}_i, \rho) = \rho^{|x_{is} - x_{ik}|}, \quad (-1 < \rho < 1). \end{aligned}$$

\mathcal{M}_2 . *Errores no correlacionados con varianza constante*

$$\begin{aligned} \boldsymbol{\varepsilon}_i &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda})) \\ \mathbf{R}(\mathbf{x}_i, \sigma^2, \boldsymbol{\lambda}) &= \sigma^2 \mathbf{I}_n \end{aligned}$$

Los vectores de parámetros estimados de los dos modelos son:

$$\begin{aligned} \mathcal{M}_1 &: \hat{\boldsymbol{\beta}}_1 = (14.880, 0.214)^T, \quad \hat{\psi}_1 = 0.042^2, \quad \hat{\sigma}_1^2 = 0.402^2, \quad \hat{\rho} = 0.935, \quad \hat{\varpi} = 0.075 \\ \mathcal{M}_2 &: \hat{\boldsymbol{\beta}}_2 = (15.029, 0.202)^T, \quad \hat{\psi}_2 = 0.207, \quad \hat{\sigma}_2^2 = 0.10^2. \end{aligned} \quad (5-29)$$

El objetivo es diseñar un experimento que permita discriminar y estimar los vectores de parámetros en los dos modelos.

Para el modelo 1 el vector de parámetros es $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1^T, \psi_1, \sigma_1^2, \rho, \varpi)^T$ y para el modelo 2 es $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2^T, \psi_2, \sigma_2^2)^T$. Adicionalmente se encontraron los diseños D-óptimos para cada modelo y el diseño T_W^I -óptimo con el fin de tener un panorama general de los diseños individualmente. Los datos del estudio inicial son únicamente utilizados para obtener buenos vectores de parámetros locales.

Con fines ilustrativos, se tomó a \mathcal{X} como el conjunto de tiempos de medición del estudio en el periodo de 2 a 24 horas

$$\mathcal{X} = \{2, 4, 6, 8, 10, 24\} \text{ horas} \quad (5-30)$$

y se fijaron dos muestras por individuo, es decir, $n = 2$. Así, la región de diseño $\bar{\mathcal{X}}^2$ contiene $C_{6,2} = 15$ combinaciones posibles de secuencias de elementos tomados de \mathcal{X} :

$$\bar{\mathcal{X}}^2 = \{(x_1, x_2) : x_s \in \mathcal{X}\}. \quad (5-31)$$

Diseños D-óptimos y T_W^I -óptimos locales

A continuación se muestran los diseños D-óptimos para cada modelo y el diseño T_W^I -óptimo para discriminación.

$$\zeta_{D_1}^* = \left\{ \begin{array}{ccc} (2, 6) & (2, 24) & (6, 24) \\ 0.63 & 0.14 & 0.23 \end{array} \right\} \quad (5-32)$$

$$\zeta_{D_2}^* = \left\{ \begin{array}{cc} (2, 6) & (2, 8) \\ 0.42 & 0.58 \end{array} \right\} \quad (5-33)$$

$$\zeta_{T_W^I}^* = \left\{ \begin{array}{cc} (2, 24) & (8, 10) \\ 0.60 & 0.40 \end{array} \right\} \quad (5-34)$$

Se observa que para el modelo 1 el diseño D-óptimo permite estimar θ_1 en forma óptima si aproximadamente el 63% de los pacientes son observados en 2 y 6 horas después de la administración del medicamento, 14% en (2,24) horas, y 23% en (6,24) horas. Para el modelo 2 el diseño es de dos puntos y la mayoría de los pacientes deben ser observados en 2 y 8 horas después de la administración. El diseño para discriminación es un diseño de dos puntos donde la mayoría de los pacientes se deben observar inmediatamente después de la administración y la segunda muestra al final del periodo de tiempo.

Con el teorema de equivalencia respectivo, gráficamente se verifica que en efecto los diseños encontrados son $\Phi_j(\omega)$ -óptimos, ver Figuras 5-2,5-3 y 5-4

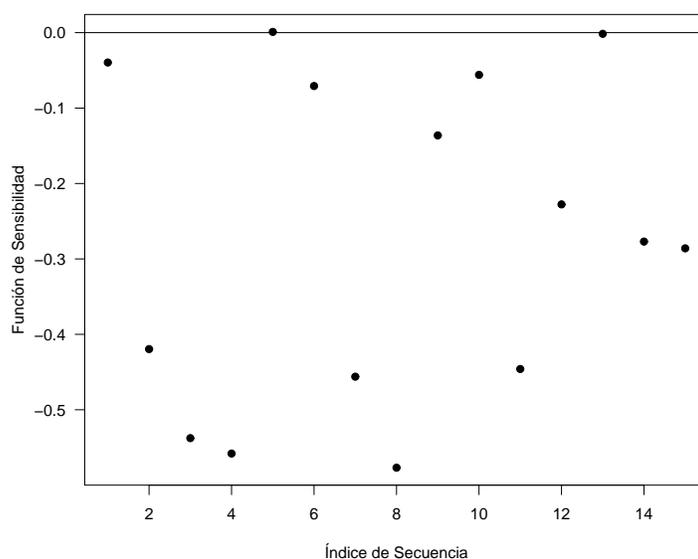


Figura 5-2.: Diseño D-óptimo local Modelo 1.

Diseños DT_W^α -óptimos locales

El diseño óptimo compuesto DT_W^α se construyó para tres valores de α : 0.25, 0.50 y 0.75. Los diseños se calcularon optimizando el criterio $\Phi_\alpha^{DT_W}(\omega)$ a través de un algoritmo implementado en el software estadístico R [41]. La función `nlminb` fue utilizada en la optimización sobre la región $\bar{\mathcal{X}}^3$. Los diseños DT_W^α -óptimos locales para los valores de α considerados son los siguientes:

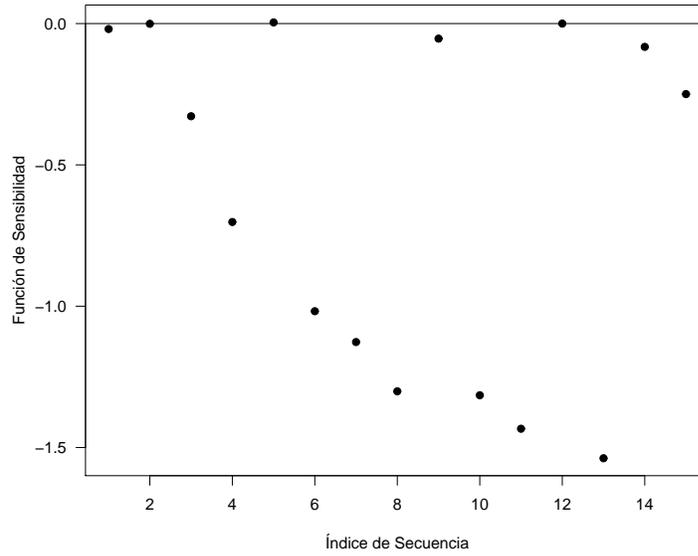


Figura 5-3.: Diseño D-óptimo local Modelo 2.

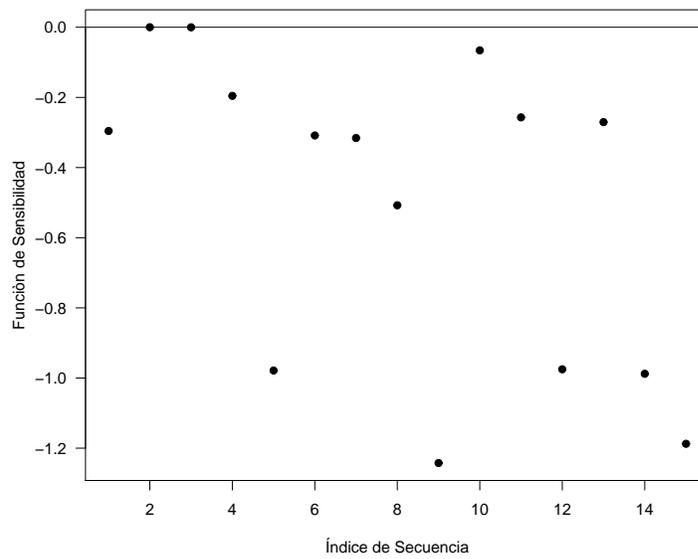


Figura 5-4.: Diseño T_W^I -óptimo local.

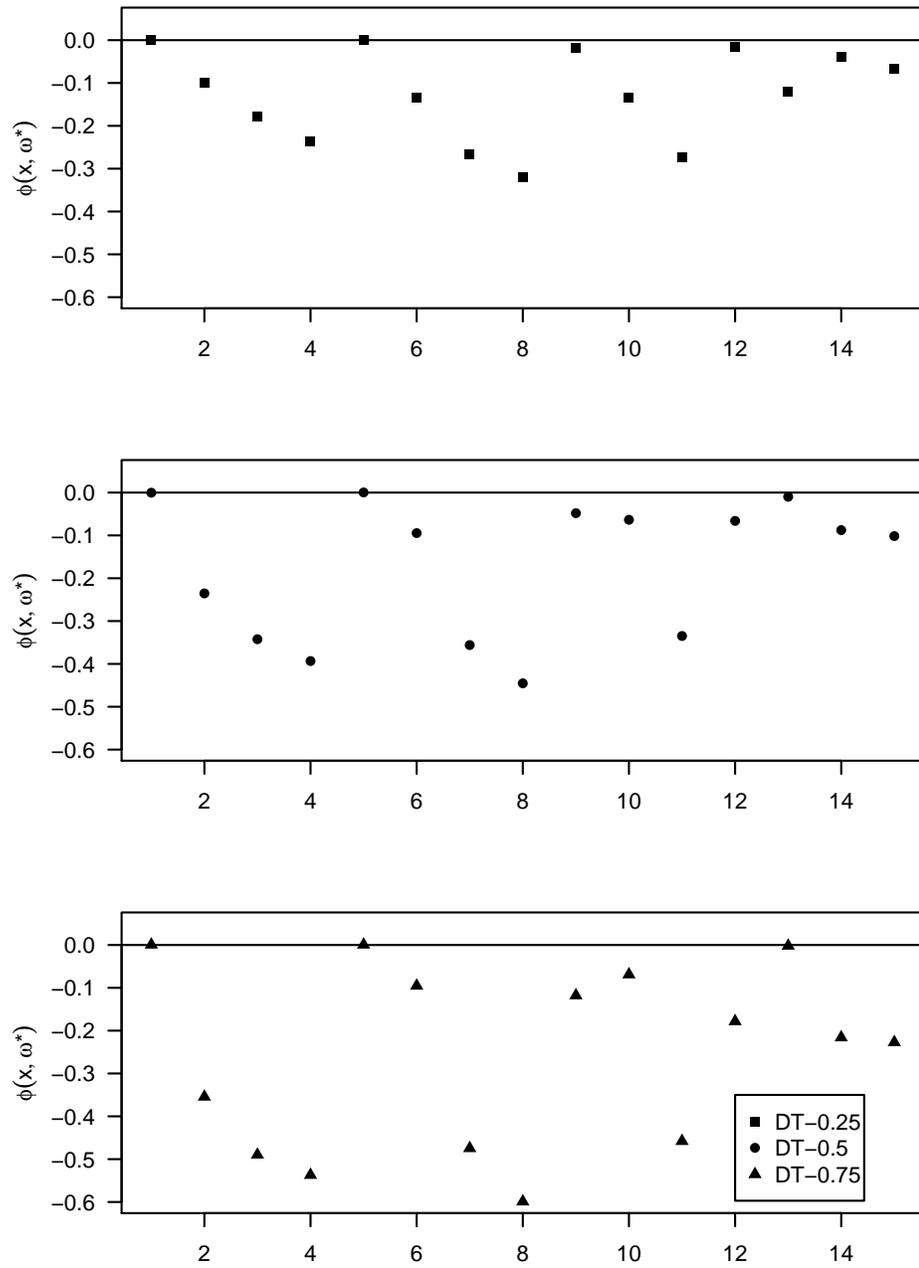


Figura 5-5.: Diseños DT_W^α -óptimos locales.

$$\zeta_{DT_W^{0.25}}^* = \left\{ \begin{array}{ccc} (2, 4) & (2, 24) & (6, 8) \\ 0.31 & 0.46 & 0.23 \end{array} \right\} \quad (5-35)$$

$$\zeta_{DT_W^{0.50}}^* = \left\{ \begin{array}{ccc} (2, 4) & (2, 24) & (8, 10) \\ 0.18 & 0.54 & 0.28 \end{array} \right\} \quad (5-36)$$

$$\zeta_{DT_W^{0.75}}^* = \left\{ \begin{array}{ccc} (2, 4) & (2, 24) & (8, 10) \\ 0.07 & 0.58 & 0.35 \end{array} \right\} \quad (5-37)$$

Los tres diseños son de tres puntos y los diseños para $\alpha = 0.5$ y $\alpha = 0.75$ coinciden en los puntos de soporte pero difieren en el peso asignado al primer punto. En la Figura 5-5 se tiene el gráfico de la derivada direccional del criterio en la dirección del diseño unitario en \mathbf{x} . Se verifica que los diseños encontrados son óptimos.

Comparación de los criterios de optimalidad

Además de los diseños óptimos obtenidos con los criterios para estimación y discriminación, se encontró el diseño poblacional ζ_P^* usando el criterio producto sugerido por Waterhouse *et al.* (2005). Con este diseño las secuencias de muestreo son (2, 6), (2, 24) y (6, 24) con frecuencias respectivas: 0.77, 0.11 y 0.12.

Tabla 5-1.: D-eficiencias y T_W^I -eficiencias de diseños poblacionales

Diseño	$D_1 - \text{ef}(\zeta)$	$D_2 - \text{ef}(\zeta)$	$T_W^I - \text{ef}(\zeta)$
$\zeta_{D_1}^*$	1	0.9023	0.4891
$\zeta_{D_2}^*$	0.7351	1	0.1139
$\zeta_{T_W^I}^*$	0.7799	0.8056	1
ζ_P^*	0.9817	0.9403	0.3623
$\zeta_{DT_W^{0.25}}^*$	0.9214	0.8406	0.8920
$\zeta_{DT_W^{0.50}}^*$	0.8591	0.8164	0.9734
$\zeta_{DT_W^{0.75}}^*$	0.8190	0.8113	0.9939

En la Tabla 5-1 se muestran las D-eficiencias y T_W^I -eficiencias para los diseños D-, T_W^I -, DT_W^α -óptimos y el diseño producto.

Se puede observar que los diseños D-óptimos tienen eficiencias por debajo del 50% para discriminar entre los modelos, el diseño T_W^I tiene eficiencias alrededor del 79% para estimación y aunque el diseño producto ζ_P^* tiene un desempeño alto para estimación su eficiencia es baja con respecto al diseño T_W^I para discriminación. En cuanto a los diseños compuestos, el valor de $\alpha = 0.25$ proporciona eficiencias relativamente similares para estimación y discriminación.

En la Figura 5-6 se representan gráficamente las eficiencias para un rango de diseños compuestos incluyendo el diseño producto ζ_P^* ($\alpha = 0$) y el diseño T_W^I -óptimo ($\alpha = 1$). Se observa

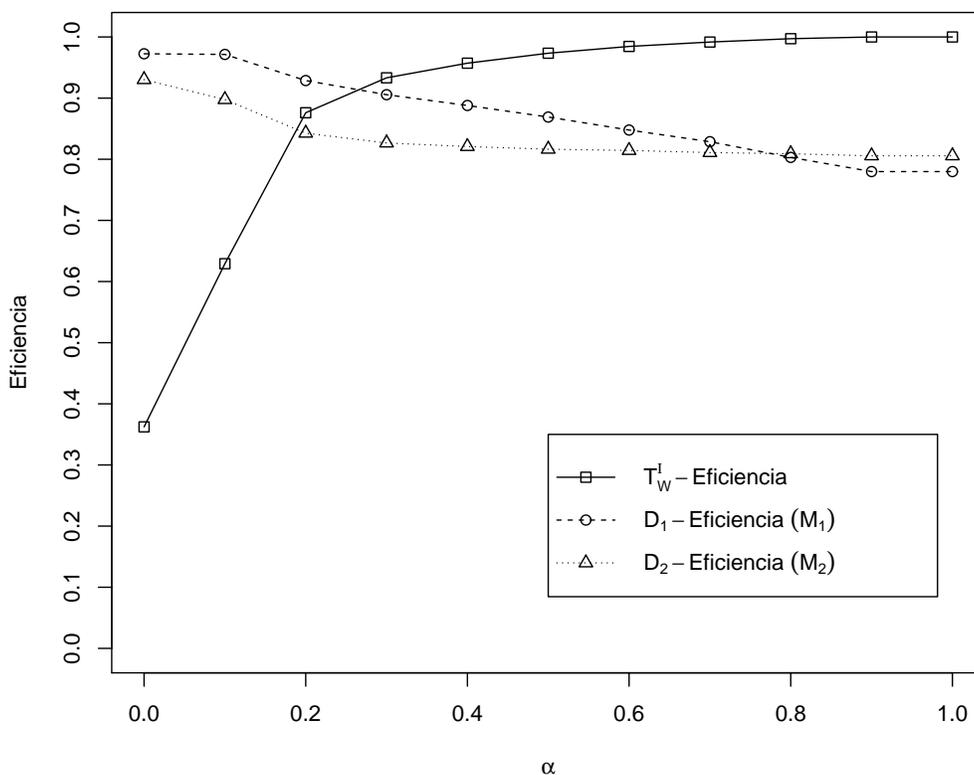


Figura 5-6.: T_W^I -eficiencias y D-eficiencias de los diseños DT_W^α -óptimos locales.

claramente un crecimiento de la T_W^I -eficiencia a medida que aumenta la importancia para discriminar entre los modelos, esto es cuando el valor de α aumenta, y un crecimiento en las D-eficiencias cuando aumenta el interés en la estimación, es decir, cuando α disminuye. Un valor de α que proporcionaría un buen compromiso entre los dos objetivos, estimación y discriminación, estaría alrededor de $\alpha = 0.27$ con una eficiencia levemente inferior para estimación en el modelo reducido.

6. Conclusiones y recomendaciones

6.1. Conclusiones

En este trabajo se construyó un criterio de optimalidad con dos objetivos simultáneos: estimación de parámetros y discriminación entre dos modelos no lineales de efectos mixtos anidados, con matrices de covarianzas intra-individuos diferentes. Bajo el enfoque de la teoría de diseños óptimos aproximados se estableció y se demostró el respectivo teorema de equivalencia. Este resultado fue utilizado para verificar la optimalidad de los diseños encontrados. El análisis del criterio se realizó a partir de una consideración particular con respecto a la ponderación asignada a los criterios de optimalidad individuales. No obstante el uso de otras ponderaciones conlleva igualmente a la formulación y demostración del respectivo teorema de equivalencia. La metodología propuesta se ilustró mediante un ejemplo. Para calcular los diseños se desarrollaron programas en el software estadístico R.

Para resolver el problema de discriminación entre modelos se propuso una extensión del criterio de T-optimalidad. El criterio se construyó para dos clases de modelos mixtos: marginales tipo I y marginales tipo II, y en cada caso se presentó la justificación estadística del criterio. Se enunció y se demostró el teorema de equivalencia para cada criterio, utilizado para verificar la optimalidad de los diseños. Además se mostró cómo se puede aplicar el criterio propuesto para modelos marginales tipo II a una clase más general de modelos no lineales de efectos mixtos.

Por medio de cuatro ejemplos de estudios reales se ilustraron las metodologías propuestas. Se demostró que los diseños propuestos tienden a superar a los diseños obtenidos bajo otros enfoques. Para los estudios de aplicación considerados, de acuerdo con los resultados obtenidos, en general se observa que los diseños óptimos asignan un peso considerablemente mayor a un solo punto. Esto conduce a la posibilidad de diseños de un solo grupo con un número pequeño de muestras por individuo. Para calcular los diseños se desarrollaron programas en el software estadístico R.

Finalmente, las metodologías propuestas permiten construir diseños poblacionales óptimos para modelos mixtos anidados con respecto a la matriz de covarianzas intra-individuos y grupos de individuos con diferentes diseños.

6.2. Recomendaciones

A partir del trabajo de investigación se proponen las siguientes línea de investigación.

- Diseños óptimos con múltiples objetivos. Se necesita hacer investigación en el desarrollo de metodologías eficientes que permitan la construcción de diseños óptimos con múltiples objetivos, como por ejemplo, además de estimación de parámetros y discriminación, considerar la estimación de ciertas funciones no lineales de los parámetros en estudios poblacionales.
- Diseños óptimos para modelos con múltiples respuestas y múltiples objetivos. Se necesita hacer investigación en el caso de modelos no lineales de efectos mixtos con múltiples respuestas y el problema de construir diseños para discriminación y estimación de funciones no lineales de los parámetros. En Kuczewski *et al.* (2008) se estudian diseños óptimos para discriminación entre modelos con múltiples respuestas para diseños de un solo grupo y en Waterhouse *et al.* (2005) la metodología propuesta es el criterio producto.
- Diseños óptimos para modelos no lineales con múltiples objetivos bajo el enfoque Bayesiano. Se necesita hacer investigación en la construcción de diseños óptimos bajo este enfoque. En Han y Chaloner (2004) se construyen diseños D-óptimos con diferentes a priori. Además de desarrollos numéricos para el cálculo de los diseños.
- Diseños óptimos para modelos no lineales más generales. Se propone extender la metodología desarrollada a modelos de efectos mixtos más generales incorporando características de los individuos (covariables) que podrían cambiar a través del tiempo. En esta clase de modelos la segunda etapa del modelo jerárquico se puede escribir como $\beta_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{b}_i$ donde \mathbf{A}_i es una matriz de diseño que depende de los valores de las características a nivel individual. Adicionalmente se podría considerar el caso en el que la matriz de covarianzas entre-individuos también dependa de dichas características. En Schmelter (2007a) se consideran diseños D-óptimos para modelos donde el vector de parámetros individual depende linealmente del vector de parámetros poblacionales más la componente aleatoria. En Bogacka *et al.* (2015) se consideran diseños D-óptimos para casos donde las covariables son variables de diseño combinadas con otras variables explicativas.
- Se necesita hacer investigación en métodos para hallar la constante de ponderación en el criterio compuesto para modelos no lineales de efectos mixtos.
- Es necesario realizar un estudio acerca de ciertas consideraciones adicionales con respecto a los diseños para discriminación en modelos no lineales mixtos como número de muestras por individuos y número de grupos.

A. Anexo: Conjuntos convexos y funciones convexas

En este anexo se presentan algunas propiedades de los conjuntos convexos y funciones convexas fundamentales en la solución de problemas extremos. Resultados sobre la estructura de un conjunto convexo así como propiedades de diferenciabilidad de funciones convexas son algunos de los tópicos tratados a continuación. El contenido está basado principalmente en los libros de Rockafellar (1970), Pshenichny (1971) y Uciński (2005).

A.1. Conjuntos convexos

Definición A.1.1. *Un subconjunto A de \mathbb{R}^n es llamado convexo si para todo par de puntos \mathbf{x} , \mathbf{y} en A se tiene que:*

$$\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} \in A, \quad \forall \lambda \in [0, 1].$$

Es decir, A es un conjunto convexo si cualquier segmento que une dos puntos cualesquiera del conjunto siempre pertenece al conjunto.

Definición A.1.2. *Sea A un subconjunto de \mathbb{R}^n . Una combinación convexa de elementos $\mathbf{x}_1, \dots, \mathbf{x}_\ell \in A$ es un vector de la forma $\sum_{i=1}^{\ell} \lambda_i \mathbf{x}_i$, donde $\lambda_i \geq 0$, $i = 1, \dots, \ell$ y $\sum_{i=1}^{\ell} \lambda_i = 1$.*

Definición A.1.3. *Sea A un subconjunto de \mathbb{R}^n . La cubierta convexa de A , denotada por $\text{conv}(A)$, es el conjunto de todas las combinaciones de elementos de A :*

$$\text{conv}(A) = \left\{ \sum_{i=1}^{\ell} \lambda_i \mathbf{x}_i : \mathbf{x}_i \in A, \lambda_i \geq 0, i = 1, \dots, \ell, \sum_{i=1}^{\ell} \lambda_i = 1 \right\}. \quad (\text{A-1})$$

Se puede demostrar que $\text{conv}(A)$ es el conjunto convexo más pequeño que contiene a A . Además si A es un conjunto compacto entonces $\text{conv}(A)$ es compacto.

El siguiente es uno de los resultados fundamentales sobre la generación de $\text{conv}(A)$.

Teorema A.1.1. (Teorema de Carathéodory). *Si A es un subconjunto de \mathbb{R}^n , entonces todo elemento \mathbf{x} en $\text{conv}(A)$ se puede expresar como una combinación convexa de a lo sumo $n + 1$ puntos de A .*

A.2. Funciones convexas

Definición A.2.1. Sea $f : A \rightarrow \mathbb{R}$ una función definida sobre un conjunto convexo A en \mathbb{R}^n . Entonces:

(i) Se dice que f es convexa si

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in A, \quad \forall \lambda \in [0, 1]. \quad (\text{A-2})$$

(ii) Se dice que f es estrictamente convexa si

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in A, \quad \mathbf{x} \neq \mathbf{y} \quad \forall \lambda \in [0, 1]. \quad (\text{A-3})$$

(iii) Se dice que f es cóncava (estrictamente) si $g \equiv -f$ es convexa (estrictamente).

Nótese que en la definición anterior es esencial que A sea un conjunto convexo. Las siguientes son algunas propiedades de las funciones convexas.

Teorema A.2.1. (i) Sean f y g dos funciones convexas definidas sobre un conjunto convexo A en \mathbb{R}^n . Entonces $\alpha f + \beta g$ es convexa si $\alpha \geq 0$ y $\beta \geq 0$.

(ii) Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa. Entonces f es continua.

A.3. Diferenciabilidad

En esta sección se presentan algunos resultados sobre diferenciabilidad de funciones convexas reales definidas sobre \mathbb{R}^n . La siguiente es la definición de una función diferenciable.

Definición A.3.1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Entonces f es diferenciable en un punto $\mathbf{x} \in \mathbb{R}^n$ cuando existe un vector \mathbf{x}^* tal que, para todo $\mathbf{y} \in \mathbb{R}^n$ se verifica

$$f(\mathbf{y}) = f(\mathbf{x}) + \langle \mathbf{x}^*, \mathbf{y} - \mathbf{x} \rangle + o(\|\mathbf{y} - \mathbf{x}\|), \quad (\text{A-4})$$

donde $\langle \cdot, \cdot \rangle$ es el producto interno Euclidiano y $\|\cdot\|$ es la norma Euclidiana.

Si f es diferenciable, entonces el vector \mathbf{x}^* es llamado el gradiente de f en \mathbf{x} el cual es denotado usualmente por $\nabla f(\mathbf{x})$ y

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right). \quad (\text{A-5})$$

A continuación se introduce la definición de derivada de Gâteaux.

Definición A.3.2. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Entonces la derivada de Gâteaux de f en \mathbf{x} en la dirección de \mathbf{y} , denotada por $G_f(\mathbf{x}, \mathbf{y})$, se define como

$$G_f(\mathbf{x}, \mathbf{y}) = \lim_{\alpha \rightarrow 0} \frac{f(\mathbf{x} + \alpha \mathbf{y}) - f(\mathbf{x})}{\alpha}. \quad (\text{A-6})$$

Si f es diferenciable en \mathbf{x} , entonces $G_f(\mathbf{x}, \mathbf{y})$ existe y

$$G_f(\mathbf{x}, \mathbf{y}) = \langle \nabla f(\mathbf{x}), \mathbf{y} \rangle = \sum y_i \frac{\partial f(\mathbf{x})}{\partial x_i}. \quad (\text{A-7})$$

Luego, a partir de la definición de $\nabla f(\mathbf{x})$ se tiene que

$$f(\mathbf{y}) - f(\mathbf{x}) = G_f(\mathbf{x}, \mathbf{y} - \mathbf{x}) + o(\|\mathbf{y} - \mathbf{x}\|) \quad (\text{A-8})$$

y $G_f(\mathbf{x}, \mathbf{y} - \mathbf{x})$ es llamada la derivada de Fréchet de f en \mathbf{x} en la dirección de \mathbf{y} .

Definición A.3.3. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Entonces la derivada de Fréchet de f en \mathbf{x} en la dirección de \mathbf{y} , denotada por $F_f(\mathbf{x}, \mathbf{y})$ está definida por

$$F_f(\mathbf{x}, \mathbf{y}) = \lim_{\alpha \rightarrow 0} \frac{f((1 - \alpha)\mathbf{x} + \alpha\mathbf{y}) - f(\mathbf{x})}{\alpha}. \quad (\text{A-9})$$

Nótese que si f es diferenciable entonces

$$F_f(\mathbf{x}, \mathbf{y}) = \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \quad (\text{A-10})$$

Definición A.3.4. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Entonces la derivada direccional unilateral de f en \mathbf{x} en la dirección de \mathbf{y} , denotada por $\partial f(\mathbf{x}, \mathbf{y})$, está definida por

$$\partial f(\mathbf{x}, \mathbf{y}) = \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha\mathbf{y}) - f(\mathbf{x})}{\alpha}. \quad (\text{A-11})$$

Si f es Gâteaux diferenciable en \mathbf{x} entonces la derivada direccional unilateral de f en \mathbf{x} existe en cualquier dirección y

$$\partial f(\mathbf{x}, \mathbf{y}) = G_f(\mathbf{x}, \mathbf{y}), \quad \forall \mathbf{y} \in \mathbb{R}^n. \quad (\text{A-12})$$

Las siguientes son algunas propiedades importantes sobre diferenciability de funciones convexas.

Teorema A.3.1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa y sea A un conjunto convexo, cerrado y acotado de \mathbb{R}^n . Entonces existe un punto $\mathbf{x}^* \in A$ tal que

$$f(\mathbf{x}^*) = \inf_{\mathbf{x} \in A} f(\mathbf{x}). \quad (\text{A-13})$$

Teorema A.3.2. Sea A un conjunto convexo de \mathbb{R}^n y sea $f : A \rightarrow \mathbb{R}$ una función convexa. Entonces cualquier mínimo local de f es también un mínimo global. Si f es estrictamente convexa, entonces existe a lo sumo un mínimo global.

Teorema A.3.3. Sea A un conjunto convexo de \mathbb{R}^n y sea $f : A \rightarrow \mathbb{R}$ una función convexa cuya derivada direccional unilateral $\partial f(\mathbf{x}^*, \mathbf{x} - \mathbf{x}^*)$ en $\mathbf{x}^* \in A$ existe para todo $\mathbf{x} \in A$. Entonces una condición necesaria y suficiente para que \mathbf{x}^* sea un mínimo de f está dada por:

$$\partial f(\mathbf{x}^*, \mathbf{x} - \mathbf{x}^*) \geq 0, \quad \forall \mathbf{x} \in A. \quad (\text{A-14})$$

En el siguiente teorema se establece la propiedad de diferenciabilidad direccional para funciones de la forma

$$f(\mathbf{x}) = \max_{\mathbf{z} \in Z} \varphi(\mathbf{x}, \mathbf{z})$$

.

Teorema A.3.4. *Sea $Z \subset \mathbb{R}^m$ un conjunto compacto y sea $\varphi : \mathbb{R}^n \times Z \rightarrow \mathbb{R}$. Si para $\mathbf{x}_0 \in \mathbb{R}^n$, la función $\varphi(\mathbf{x}_0 + \lambda \boldsymbol{\delta}, \mathbf{z})$ es una función continua en λ y \mathbf{z} , y además $\frac{\partial \varphi(\mathbf{x}_0 + \lambda \boldsymbol{\delta}, \mathbf{z})}{\partial \lambda}$ existe y es continua en $\lambda \in [0, 1]$ y $\mathbf{z} \in Z$, entonces f es diferenciable en \mathbf{x}_0 en la dirección de $\boldsymbol{\delta}$ y*

$$\partial f(\mathbf{x}_0, \boldsymbol{\delta}) = \max_{\mathbf{z} \in Z(\mathbf{x}_0)} \partial \varphi(\mathbf{x}_0, \mathbf{z}), \tag{A-15}$$

donde $Z(\mathbf{x}) = \{\mathbf{z} \in Z : f(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{z})\}$.

B. Anexo: Código fuente en R para el cálculo de diseños T_{W}^I – *óptimos*

#1. FUNCIONES PARA CALCULAR MATRICES DE COVARIANZA

```
library(MASS)
```

```
Mat.Inf<-function(xi,fx,th,S)
{
  x <- t(as.matrix(xi))
  X <- sapply(x,fx,th)
  Mat <- X%*%ginv(S)%*%t(X)
  return(Mat)
}
```

#2. Matriz de varianzas-covarianzas para los errores.
Matriz de Correlacion.

```
Cov.Exp<-function(xi,phi)
{
  S<-matrix(0,ncol=length(xi),nrow=length(xi))
  for(i in 1:length(xi))
  {
    for(j in 1:length(xi))
    {
      S[i,j]=phi^(abs(xi[i]-xi[j]))
    }
  }
  return(S)
}
```

Matriz de varianzas.

```
Cov.Var<-function(xi,a)
{
```

```

S<-matrix(0,ncol=length(xi),nrow=length(xi))
for(i in 1:length(xi))
{
S[i,i]=exp(-a*xi[i])
}
return(S)
}

```

#4. Matriz de varianzas-covarianzas para los errores

```

Cov.R<- function(xi,phi,sig)
{
R<-(sig^2)*Cov.Exp(xi,phi)
return(R)
}

```

#5. Matriz de covarianzas para el Modelo FULL

```

Cov.Full<-function(xi,phi,sig,d)
{
uno<-c(rep(1,length(xi)))
V <- (sig^2)*(d^2)*uno%*%t(uno)+Cov.R(xi,phi,sig)
return(V)
}

```

#6. Matriz de covarianzas modelo Reducido

```

Cov.Red<-function(xi,sig2,d2)
{
T<-Cov.Full(xi,0,sig2,d2)
return(T)
}

```

#7. Funciones para calcular el diseño Tw-\'optimo

```

festth<-function(xi,p,phi,sig,d,r)
{
t = xi
u =p[1:2]
W1 = Cov.Full(xi,phi,sig,d)
dm = exp(p[2])
IW2 = ginv(Cov.Red(xi,exp(p[1]),dm))
M = IW2%*%W1
psi = sum(diag(M))-log(det(M))-r
}

```

```

return(psi)
}

gth<-function(p,wei,phi,sig,d,r)
{
X<-apply(XI,2,festth,p,phi,sig,d,r)
w<-wei*(1/sum(wei))
G<-sum(w*X)
return(G)
}

Delta2thCor<-function(wei,phi,sig,d,p2,r)
{
out = nlminb(p2,gth,lower=c(rep(-Inf,2)),upper=c(rep(Inf,2)),
wei=wei,phi=phi,sig=sig,d=d,r=r)
p <- out$par
  w = wei
  Del<-gth(p,w,phi,sig,d,r)
return(-log(Del))
}

resTOP<-nlminb(w0,Delta2thCor,lower=c(rep(0,NG)),upper=c(rep(1,NG)),
phi=phi,sig=sig,d=d,p2=p2,r=r)

disfin<-matrix(resTOP$par,1)
nc4<-ncol(disfin)
pesos4<-disfin*(1/sum(disfin))
pesos4
ToptP<-rbind(XI,pesos4)

#8. Funciones para calcular el parametro p2

FT<-function(xi,phi,a,sig,d,p,r)
{
W1 = Cov.Full(xi,phi,a,sig,d)
dm = exp(p[2])
IW2 = ginv(Cov.Red(xi,exp(p[1]),dm))
M = IW2*%*%W1
psi = sum(diag(M))-log(det(M))-r
return(psi)
}

```

```
Thopt<-function(p,phi,a,sig,d,zeta0,w0,r)
{
X<-as.matrix(zeta0)
Q<-apply(X,2,FT,phi,a,sig,d,p,r)
Fin<-sum(w0*Q)
return(log(Fin))
}

Th2Opt<-nlminb(p2,Thopt,lower=c(rep(-Inf,length(p2))),upper=c(rep(Inf,length(p2))),
,phi=phi,a=a,sig=sig,d=d,zeta0=zeta0,w0=w0,r=r);
tens=Th2Opt$par; Tx<-apply(XI,2,FT,phi,a,sig,d,tens,r)
```

Bibliografía

- [1] A. C. Atkinson, “DT-optimum designs for model discrimination and parameter estimation”, *Journal of Statistical Planning and Inference*, vol. 138, pp. 56-64, 2008.
- [2] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*, Oxford Science Publications, New York, 1992.
- [3] A. C. Atkinson and V. V. Fedorov, “The design of experiments for discriminating between two rival models”, *Biometrika*, vol. 62, pp. 57-70, 1975.
- [4] S. L. Beal and L. B. Sheiner, “Estimating population kinetics”, *CRC Critical Reviews in Biomedical Engineering*, vol. 8, pp. 195-292, 1982.
- [5] B. Bogacka, A. H. M. Mahbub Latif, S. Gilmour and K. Youdim, “Optimum designs for non-linear mixed effects models in the presence of covariates”, *Biometrics*, 2015.
- [6] K. Bollen, *Estructural Equations with Latents Variables*, John Wiley & Sons, New York, Inc, 1989.
- [7] M. E. Castañeda and V. I. López-Ríos, “Optimal population designs for discrimination between two nested nonlinear mixed effects models”, *Revista Ciencia en Desarrollo*, vol. 7, pp. 71-81, 2016.
- [8] H. Chernoff, “Locally optimal designs for estimating parameters”, *The Annals of Mathematical of Statistics*, vol. 24, pp. 586-602, 1953.
- [9] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman & Hall, London, 1974.
- [10] M. Crowder and J. Tredger, “The use of exponentially damped polynomials for biological recovery data ”, *Applied Statistics*, vol. 32, pp. 15-18, 1981.
- [11] M. Davidian and D. M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall, London, 1995.
- [12] L. K. Debusho, *Optimal designs for linear mixed models*, Tesis, Sudáfrica, 2004.
- [13] H. Dette, A. Pepelyshev and T. Holland-Letz “Optimal designs for random effect models with correlated errors with applications in population pharmacokinetics”, *The Annals of Applied Statistics*, vol. 4, pp. 1430-1450, 2010.

-
- [14] E. Demidenko, *Mixed Models: Theory and Applications*, John Wiley & Sons Inc, New York, 2004.
- [15] V. V. Fedorov, *Theory of Optimal Experiments*, Academic Press, New York, 1972.
- [16] V. V. Fedorov and P. Hackl, *Model-Oriented Design of Experiments*, Springer, New York, 1997.
- [17] V. V. Fedorov and S. Leonov, "Population pharmacokinetic measures, their estimation and selection of sampling times", *Journal of Biopharmaceutical Statistics*, vol. 17, pp. 919-941, 2007.
- [18] R. Gagnon and S. Leonov, "Optimal population designs for PK models with serial sampling", *Journal of Biopharmaceutical Statistics*, vol. 15, pp. 143-163, 2005.
- [19] C. Han and K. Chaloner, "Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics", *Biometrics*, vol. 60, pp. 25-33, 2004.
- [20] D. Hand and M. Crowder, *Practical Longitudinal Data Analysis*, Chapman & Hall, London, 1996.
- [21] Y. Hashimoto and L. Sheiner, "Designs for population pharmacodynamics: value of pharmacokinetic data and population analysis", *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 19, pp. 333-353, 1991.
- [22] T. Holland-Letz, H. Dette and D. Renard, "Efficient algorithms for optimal designs with correlated observations in pharmacokinetics and dose-finding studies", *Biometrics*, vol. 68, pp. 138-145, 2012.
- [23] B. Jones and J. Wang, "Constructing optimal designs for fitting pharmacokinetic models", *Statistics and Computing*, vol. 9, pp. 209-218, 1999.
- [24] J. Kiefer, "Optimum Experimental Designs", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 21, pp. 272-319, 1959.
- [25] J. Kiefer and K. Wolfowitz, "The equivalence of the two extremum problems", *Canadian Journal of Mathematics*, vol. 12, pp. 363-366, 1960.
- [26] B. Kuczewski, B. Bogacka and D. Uciński, "Optimum designs for discrimination between two nonlinear multivariate dynamic mixed-effects models", *Biometrical Letters*, vol. 45, pp. 1-28, 2008.
- [27] J. Lin and B. Wei, "Testing for heteroscedasticity and/or autocorrelation in longitudinal mixed effect nonlinear models with AR(1) errors", *Communications in Statistics-Theory and Methods*, vol. 36, pp. 567-586, 2007.

-
- [28] J. López-Fidalgo, C. Tommasi and P.C. Trandafir, “An optimal experimental design criterion for discriminating between non-normal models”, *Journal of The Royal Statistical Society Series B (Statistical Methodology)*, vol. 69, pp. 231-242, 2007.
- [29] V. I. López-Ríos, Diseños óptimos para discriminación y estimación en modelos no lineales, Ph. D Tesis, México, 2008.
- [30] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York, 1988.
- [31] F. Mentré, P. Burtin, Y. Merlé, J. Bree, A. Mallet and J. Steimer, “Sparse-sampling optimal designs in pharmacokinetics and toxicokinetics”, *Drug Information Journal*, vol. 29, pp. 997-1019, 1995.
- [32] F. Mentré, A. Mallet and D. Baccar, “Optimal design in random-effects regression models”, *Biometrika*, vol. 84, pp. 429-442, 1997.
- [33] T. Mielke, “Approximation of the Fisher information and design in nonlinear mixed effects models”, in *Experiments for Processes with Time or Space Dynamics*, Isaac Newton Institute for Mathematical Sciences, 2011.
- [34] L. Nie, “Optimal designs for linear mixed-effects models”, *Journal of Mathematical Science* vol. 1, 2007.
- [35] J. Nyberg, R. Hglund, M. Bergstrand, M. Karlsson and A. Hooker, “Serial correlation in optimal design for nonlinear mixed effects models”, *J Pharmacokinetic Pharmacodyn*, vol. 39, pp. 239-249, 2012.
- [36] M. Patan and B. Bogacka, “Optimum group designs for random-effects nonlinear dynamic processes”, *Chemometrics and Intelligent Laboratory Systems*, vol. 101, pp. 73-86, 2010.
- [37] J. C. Pinheiro and D. M. Bates, *Mixed-Effects Models in S and S-PLUS*, Springer, New York, 2000.
- [38] B. N. Pshenichny, *Necessary Conditions for an Extremum*, Marcel Dekker, New York, 1971.
- [39] F. Pukelsheim, *Optimal Design of Experiments*, John Wiley & Sons Inc, New York, 1993.
- [40] F. Pukelsheim and J. L. Rosenberger, “Experimental designs for model discrimination”, *Journal of the American Statistical Association*, vol. 88, pp. 642-649, 1993.

- [41] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014, ISBN 3-900051-07-0, [Online], Available: <http://www.R-project.org>.
- [42] A. Racine, A. P. Grieve, H. Fluhler and A. F. M. Smith, Bayesian Methods in Practice: Experiences in the Pharmaceutical Industry, *Bayesian Methods in Practice: Experiences in the Pharmaceutical Industry*, vol. 35, pp. 93-150, 1986.
- [43] S. Retout, E. Comets, A. Samson and F. Mentré, “Design in nonlinear mixed effects models: optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates”, *Statistics in Medicine*, vol. 26, pp. 5162-5179, 2007.
- [44] S. Retout and F. Mentré, “Further developments of the Fisher information matrix in nonlinear mixed-effects models with evaluation in population pharmacokinetics”, *Journal of Biopharmaceutical Statistics*, vol. 13, pp. 209-227, 2003a.
- [45] S. Retout and F. Mentré, “Optimisation of individual and population designs using Splus”, *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 30, pp. 417-443, 2003b.
- [46] S. Retout, F. Mentré and R. Bruno, “Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics”, *Statistics in Medicine*, vol. 21, pp. 2623-2639, 2002.
- [47] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, New Jersey, 1970.
- [48] A. Satorra and W. E. Saris, “Power of the likelihood ratio test in covariance structure analysis”, *Psychometrika*, vol. 50, pp. 83-90, 1985.
- [49] T. Schmelter, “Considerations on group-wise identical designs for linear mixed models”, *Journal of Statistical Planning and Inference*, vol. 137, pp. 4003-4010, 2007a.
- [50] T. Schmelter, “The optimality of single group designs for certain mixed models”, *Metrica*, vol. 65, pp. 183-193, 2007b.
- [51] S. D. Silvey, *Optimal Design*, Chapman & Hall, London, 1980.
- [52] J. H. Steiger, A. Shapiro and W. Browne, “On the multivariate asymptotic distribution of sequential Chi-Square statistics”, *Psychometrika*, vol. 50, pp. 253-264, 1985.
- [53] K. Smith, “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*, vol. 12, pp. 1-85, 1918.
- [54] J. R. Stroud, P. Müller and G. Rosner, “Optimal sampling times in population pharmacokinetic studies”, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 50, pp. 345-359, 2001.

-
- [55] T. W. Stroud, “Fixed alternatives and Wald’s formulation of the noncentral asymptotic behavior of the likelihood ratio statistic”, *Annals of Mathematical Statistics*, vol. 43, pp. 447-454, 1972.
- [56] D. Uciński, *Optimal Measurement Methods for Distributed Parameter System Identification*, CRC Press, Boca Raton, 2005.
- [57] D. Uciński and B. Bogacka, “T-optimum designs for multiresponse dynamic heteroscedastic models”, In A. Di Bucchianico, H. Ler, and H. P. Wynn, editors, mODa 7, Proc. 7th Int. Workshop on Model-Oriented Data Analysis, Heeze, The Netherlands, 2004, pp. 191-199, Heidelberg, Physica-Verlag, 2004.
- [58] D. Uciński and B. Bogacka, “T-optimum designs for discrimination between two multi-response dynamic models”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, pp. 3-18, 2005.
- [59] P. Vajjah and S. Duffull, “A generalisation of T-optimality for discriminating between competing models with an application to pharmacokinetic studies”, *Pharmaceutical Statistics*, vol. 11, pp. 503-510, 2012.
- [60] E. F. Vonesh, and V. M. Chinchilli, *Linear and Nonlinear Models for The Analysis of Repeated Measures*, Marcel Dekker, New York, 1997.
- [61] A. Wald, “On the efficient design of statistical investigations”, *The Annals of Mathematical Statistics*, vol. 14, pp. 134-140, 1943.
- [62] J. Wang, “Optimal parametric design with applications to pharmacokinetic and pharmacodynamic trials”, *Journal of Applied Statistics*, vol. 33, pp. 837-852, 2007.
- [63] T. H. Waterhouse, S. Redmann, S. B. Duffull and J. A. Eccleston, “Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients”, *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 32, pp. 521-545, 2005.
- [64] P. Whittle, “Some general points in the theory of optimal experimental design”, *Journal Royal Statistical Society. Series B (Statistical Methodology)*, vol. 35, pp. 123-130, 1973.