



CIMAT



Centro de Investigación en Matemáticas, A.C.

“DISEÑOS ÓPTIMOS
PARA EL MODELO DE
MICHAELIS-MENTEN”

T E S I S

Que para obtener el Grado de:

Maestro en Ciencias

con Orientación en

Probabilidad y Estadística

P r e s e n t a

Gabriel Alejandro

Negrón Pérez

Director de Tesis:

Dr. Rogelio Ramos Quiroga

Guanajuato, Gto., Febrero 2014

Dedicatoria

*A mi hija, Vanessa,
a mi esposa, Alexis,
a mis padres, Humberto y Beatriz,
y a mis hermanos, Gustavo, Angel y Arturo.*

Agradecimientos

Mi más profunda gratitud al Dr. Rogelio Ramos Quiroga por su guía, consejos y paciencia para la realización de este trabajo.

Le doy un agradecimiento infinito al pueblo mexicano, que a través del CONACYT solventó mis estudios de maestría, lo cual permitió que concentrara mis esfuerzos en el estudio y acabar este periodo de manera exitosa.

Agradezco a toda la comunidad del CIMAT. En especial al cuerpo académico del área de probabilidad y estadística, por compartir sus conocimientos y enseñarme varias formas de deleitarme con estas disciplinas. Además, me han enriquecido con sus consejos y experiencia que me serán útiles para mi vida futura. Jamás los olvidaré.

Agradezco cariñosamente a mis padres y hermanos, quienes me han ayudado en todo momento, desde siempre.

Mi más amorosa gratitud a mi esposa, por su apoyo incondicional, su comprensión, su tiempo y por el amor que me hace sentir.

A mi hija le doy el más enternecedor agradecimiento, pues sin importar su corta edad, supo comprenderme durante los largos momentos.

Gracias a los amigos que hice en Guanajuato, por haber convergido nuestros caminos este tiempo, por compartir sus experiencias y sus personas.

Resumen

Muchas de las reacciones en la cinética de enzimas se rigen por la ecuación de Michaelis-Menten. La caracterización de estas reacciones requiere la estimación de los parámetros de K_m y V_{max} que determinan la ecuación de Michaelis-Menten y esto se hace mediante la observación de las tasas de reacciones en un conjunto de concentraciones de sustrato. Esta elección de las concentraciones de sustrato se investigó mediante la determinación del diseño óptimo usando el criterio D-óptimo que maximiza el determinante de la matriz de información de Fisher.

La matriz de información de Fisher se formula para el modelo no lineal de Michaelis-Menten y para las distribuciones normales. Diseñar experimentos para modelos no lineales de ajuste, se complica por el hecho de que las varianzas de las estimaciones de los parámetros dependen de los valores desconocidos de estos parámetros. Se hace una revisión del modelo no lineal con distintos enfoques para el diseño óptimo y se recomienda el diseño D-óptimo local y el diseño D-óptimo bayesiano para el modelo no lineal de Michaelis-Menten.

Índice general

1. Introducción.	1
2. Introducción a la Cinética de la Enzima	6
2.1. Introducción.	6
2.2. Velocidad de Reacción.	7
2.3. Leyes de Velocidad y Orden de Reacción.	7
2.4. Catálisis.	8
2.4.1. Catálisis Enzimática y Cinética.	9
2.5. Mecanismo de Michaelis-Menten.	9
2.6. Estimación de los Parámetros del Modelo de Michaelis-Menten.	13
3. Modelos Estadísticos.	15
3.1. Modelos Lineales.	15
3.1.1. Estimación del Modelo Lineal.	16
3.2. Modelos No Lineales.	19
3.2.1. Estimación del Modelo No lineal.	21
3.3. Transformación a un Modelo Lineal.	22
3.4. Estimación de los Parámetros en un Sistema No Lineal.	25
4. Criterios de Optimalidad.	30
4.1. Introducción.	30
4.2. Teoría del Diseño Óptimo	30
4.2.1. Diseños Continuos y Diseños Exactos.	31
4.2.2. Matriz de Información para los Diseños Continuos y los Diseños Exactos.	32

4.3. Criterios de Optimalidad.	33
4.4. Funciones criterios y sus propiedades.	34
4.4.1. D-optimalidad.	35
4.4.2. A-optimalidad.	36
4.4.3. G-optimalidad.	36
4.4.4. E-optimalidad.	37
4.4.5. c-optimalidad.	37
4.4.6. L-optimalidad.	38
4.4.7. Φ_p -optimalidad.	38
4.4.8. L_p -optimalidad.	38
4.4.9. I-optimalidad.	39
4.4.10. MV-optimalidad.	39
4.4.11. Criterio minimax.	40
4.5. Eficiencia de los diseños.	42
5. Teorema de Equivalencia.	44
5.1. Teorema General de Equivalencia.	46
6. Modelo de Michaelis-Menten.	53
6.1. Introducción.	53
6.2. Modelo de Michaelis-Menten.	53
6.3. Matriz de Información de Fisher.	54
6.4. Diseños Localmente D-óptimos.	56
6.5. Diseños D-óptimos Bayesianos.	62
7. Conclusiones.	66

Capítulo 1

Introducción.

En el nivel actual de desarrollo de la ciencia y la tecnología muchas investigaciones en la física, la biología, la química, la metalurgia, etc, requieren creación de experimentos complicados y costosos. La medición de una cantidad experimental se realiza siempre bajo la influencia de algunos obstáculos que no pueden ser eliminados por completo a pesar de los esfuerzos de los investigadores para mantenerlos al mínimo. Debido a esto, el investigador no se ocupa de cantidades deterministas, pero con cantidades aleatorias. En algunos casos, las cantidades medidas son al azar, por la variedad de su naturaleza. Es necesario hacer frente a la medición de las cantidades en la mecánica cuántica, en las investigaciones biológicas, en ciertos problemas de cinética química, y otras ramas de la ciencia. La necesidad de aplicar estadística matemática a la reducción de los resultados de las mediciones es evidente cuando los componentes aleatorios son acordes con los mismos resultados. Los métodos correspondientes de reducción han sido utilizados en la práctica experimental.

Durante mucho tiempo, la atención de la estadística matemática se centró en la perfección de los métodos de reducción cuando el método de llevar a cabo el experimento fue preestablecido. La elección de la propia experiencia, es decir, cuándo y dónde llevar la medición, se determina principalmente por la intuición del experimentador. Durante este tiempo fue necesario para hacer frente a los problemas que son relativamente sencillos de puntos de vista teóricos y experimentales, y que no requieren gastos significativos, es decir, los recursos financieros, tiempo, recursos materiales limitados. Las pérdidas relativas a los errores de la solución intuitiva para el método de llevar a cabo el experimento no se cumplen muy a menudo y no eran esenciales.

El desarrollo de la ciencia y la tecnología llevó a la complicación natural en la interpretación teórica de

los resultados obtenidos, y en los métodos para llevar a cabo investigaciones experimentales necesarias. Muchas situaciones experimentales complicadas condujeron a un fuerte aumento en el costo de las investigaciones experimentales. Como un ejemplo, se pueden citar las investigaciones en el ámbito de la física de las partículas elementales donde la necesidad de la construcción de aceleradores de gran alcance hace mediciones muy caras. Por lo tanto, el problema de la extracción de una mayor cantidad de datos a partir de los procesos, en estudio con recursos finitos, es actualmente muy real. Confiar en la intuición del investigador para la solución de un problema dado, porque cada vez hay menos esperanzas. En relación con esto, es absolutamente necesario para dar una amplia clase de métodos que darían no sólo los medios de reducción de los datos experimentales, pero también permitiría la organización del experimento de la manera óptima. El aparato matemático utilizado en la organización óptima de los experimentos se basa en una composición de métodos estadísticos y los métodos de resolución de problemas matemáticos externos. Cada vez más a menudo, la estadística matemática es necesaria para la construcción racional y la elucidación de las propiedades básicas de los criterios de optimalidad de un experimento. Después, el problema de la organización óptima de un experimento conduce a la solución de algún problema extremo.

El diseño se utiliza en ese caso, cuando el experimentador está interesado en condiciones en las que el proceso que está siendo investigado satisface algunos de los criterios de optimalidad. Por ejemplo, en el desarrollo de un nuevo proceso tecnológico químico, los criterios de optimalidad requieren la producción máxima de los productos de las reacciones. En este caso, el diseño consiste en encontrar los valores de la temperatura, la presión de los reactivos, su porcentaje de concentración, y así sucesivamente, para que los requisitos establecidos sean satisfechos.

El uso cada vez mayor de técnicas de optimización y potentes ordenadores modernos da lugar a la utilización de modelos más complejos y realistas en la modelación de datos. Sin embargo, a juzgar por una revisión de la literatura en el diseño óptimo de experimentos (Silvey (1980); Atkinson et al (2007)), la mayoría de la investigación en esta área se mantiene enfocada en modelos lineales, mientras que, de hecho, los modelos no lineales son comúnmente utilizados en muchas áreas tales como la química, la farmacología y la farmacocinética. Atkinson et al. (2007), por ejemplo, dieron una excelente introducción a la teoría del diseño óptimo de experimentos.

El diseño óptimo ha sido estudiado desde tan temprano como en 1918, cuando el artículo de Smith (1918) define el objetivo de minimizar el error de predicción del peor caso en la construcción de modelos polinomiales. En este trabajo, se presenta lo que se conoce como G-optimalidad que minimiza el máximo

sobre el espacio de diseño de la varianza normalizada, tal como se define por Kiefer y Wolfowitz (1959). En el mismo artículo de Kiefer y Wolfowitz (1959), le dieron el nombre de D-optimalidad el criterio introducido por Wald (1943), que pone el énfasis en la calidad de las estimaciones de los parámetros. Un diseño D-óptimo maximiza el determinante de la matriz de información de Fisher, por lo tanto, reducir al mínimo el volumen de la confianza elipsoide de cualquier ofrece estimaciones de los parámetros. Kiefer y Wolfowitz (1959) también se relacionan estos dos criterios alfabéticos por el Teorema General de Equivalencia. Todos los documentos mencionados hasta ahora sólo tienen en cuenta los modelos que son lineales en sus parámetros. Ciertas complicaciones surgen en el diseño óptimo de experimentos cuando el modelo no es lineal en los parámetros, el más notable de los cuales es que los modelos no lineales requieren estimaciones de los parámetros del modelo que se caracterizan por un diseño óptimo que se construirá. Típicamente las estimaciones iniciales de los parámetros se basan en los resultados de los estudios anteriores o conjeturas de los expertos. Los diseños óptimos encontrados mediante estimaciones puntuales se dice que son localmente óptimos.

Aunque el trabajo en la teoría de diseño óptimo para los modelos lineales se remonta hasta 1918, los modelos no lineales no aparecen en la literatura diseño óptimo hasta 1959. Investigaciones de diseños localmente D-óptimos para modelos no lineales, en los que los autores justifican la dependencia de tales diseños antes de conocimiento de los valores de los parámetros al afirmar que "en los problemas prácticos es casi invariablemente el caso de que parte de dicha información está disponible, y esto proporcionará entonces la base de un primer diseño". La metodología de la regresión no lineal se discute en detalle por Ratkowsky (1990), Bates y Watts (1988), y Seber y Wild (2003). Un método para obtener y presentar intervalos de varios tamaños para un único parámetro en un modelo no lineal de múltiples parámetros se presenta por Cook y Weisberg (1990). Muchos de los criterios se han aplicado para el diseño de los modelos no lineales. Por ejemplo, Box y Lucas (1959), Ford et al. (1989), Haines (1993), y Hedayat et al. (2003) obtienen Diseños D-óptimos. El diseño Minimax lo estudió Dette y Sahm (1998). Dette y O'Brien (1999) proponen un nuevo criterio de optimización, que ellos llaman I-L-optimalidad. En base a la varianza predicha, este criterio es invariante con respecto a diferentes parametrizaciones del modelo. G- y D-optimalidad pueden ser vistos como sus casos especiales. El uso de E-optimalidad se discute en Dette y Haines (1994) y Dette et al. (2004).

Ratkowsky (1990) ofrece una introducción a los modelos no lineales, con énfasis en las propiedades de una variedad de modelos para uno o unos pocos factores. Algunos modelos no lineales han recibido una especial atención debido a la importancia del papel que desempeñan en algunas investigaciones

científicas. Por ejemplo, el diseño para el modelo propuesto por Michaelis y Menten (1913) se discute por Dette et al. (2003), Dette y Biedermann (2003) y Dette et al. (2005). Dette et al. (2003) investigan los diseños eficientes mediante la maximización de un mínimo de estándar E-eficiencia en el modelo de Michaelis-Menten. Se muestra en muchos casos que los diseños óptimos son compatibles en sólo dos puntos y que los puntos de apoyo y los pesos correspondientes se pueden caracterizar de forma explícita. Por otra parte, un estudio numérico indica que dos diseños de puntos son por lo general muy eficientes, incluso si no son óptimos.

El modelo de Michaelis-Menten es uno de los modelos más utilizados en las ciencias biológicas. El modelo es quizás el más comúnmente usado para estudiar la reacción enzimática que es de gran importancia en la farmacología, la biología y la investigación médica. Más específicamente, el modelo se utiliza para describir las funciones de saturación para numerosos fenómenos físicos y biológicos. Las aplicaciones específicas se pueden encontrar en Cressie y Keightley (1979), Johansen (1984), Beverton y Holt (1957), Cornish-Browden (1979), Hay, Meznarich, DiGiacomo, Hirst, Zerbe (1988), sólo por nombrar a algunos pocos. Los diseños óptimos para el modelo de Michaelis-Menten se han estudiado por numerosos autores. Dette y Biedermann (2003) sugirieron diseños D-óptimos durante un cierto intervalo para el parámetro no lineal para el modelo de Michaelis-Menten. La literatura sobre el diseño óptimo para modelos no lineales ha sido relativamente escasa debido a esta dependencia de los valores de los parámetros. Chaloner y Verdinelli (1995) y Clyde (2001), revizaron un método para evitar la necesidad de estimaciones puntuales en el diseño óptimo, es el uso de diseños experimentales bayesianos, en el que antes de las distribuciones se colocan en los parámetros del modelo. Song y Wong (1998) propusieron diseños D-óptimos bayesianos para el modelo de Michaelis-Menten. Matthews y Allcock (2004), presentaron una variedad de diseños óptimos bayesianos para el modelo de Michaelis-Menten con aplicaciones en enzimología.

En el área de la farmacocinética (PK, el estudio de la evolución temporal de la concentración cambiante de un fármaco en el cuerpo), por ejemplo, cuando se utilizan comúnmente modelos no lineales, la mayoría del trabajo en el diseño óptimo ha sido teórico, con poco en el camino de la aplicación. La literatura en la zona comienza con D'Argenio (1981), que discute la elección óptima de los tiempos de muestreo en un estudio PK, pero limita la optimización de un único sujeto.

Los objetivos principales de esta tesis son de discutir los diseños óptimos para modelos no lineales y en particular obtener diseños óptimos para el modelo de Michaelis-Menten, tanto con el criterio clásico de D-óptimalidad local como en el enfoque bayesiano, se presentan también aplicaciones a ejemplos de

la vida real que se plantean en la cinética enzimática. Aunque la aplicación de estas técnicas ha sido la novela en sí misma, esta tesis ha presentado los métodos de diseño de experimentos con el modelo no lineal.

En primer lugar nos centramos en lo que es la cinética de enzimas, la relación entre la cinética enzimática y el modelo de Michaelis-Menten y luego se discute el mecanismo del modelo de Michaelis-Menten en el capítulo 2. En el siguiente capítulo, Capítulo 3, se describe el modelo lineal y el modelo no lineal, se describirá el procedimiento de estimación de los parámetros. Se presenta una manera mucho más fácil de estimar los parámetros de las estimaciones encontradas utilizando el modelo lineal.

En el capítulo 4 se representa la teoría del diseño óptimo, diversos criterios sobre el diseño óptimo.

En el capítulo 5 se hace una extensa revisión de la teoría sobre el teorema de equivalencia, así como los puntos de soporte y los pesos del diseño.

En el capítulo 6 se discute el diseño óptimo para el modelo de Michaelis-Menten.

Capítulo 2

Introducción a la Cinética de la Enzima

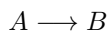
2.1. Introducción.

Cinética es el estudio de los tipos de procesos químicos en un esfuerzo por entender qué es lo que influye en estos tipos y para desarrollar las teorías que se pueden usar para predecir ellos. El conocimiento de las velocidades de reacción tiene muchas aplicaciones prácticas, por ejemplo en el diseño de un proceso industrial, en la comprensión de la compleja dinámica de la atmósfera y en la comprensión de la compleja interacción de las reacciones químicas que son la base de la vida, etc.

En un nivel más fundamental, queremos entender lo que ocurre con las moléculas en una reacción química, que es lo que sucede en un solo encuentro reactivo entre dos moléculas reactivas. La mayoría de las reacciones son reacciones espontáneas. Estas reacción se produce de izquierda a derecha hasta que todos los reactivos se convierten en productos. Una reacción puede ser lento o puede ser rápido. Al comprender esto podemos ser capaces de desarrollar teorías que pueden ser utilizados para predecir el resultado y la velocidad de las reacciones.

2.2. Velocidad de Reacción.

La velocidad de una reacción dice en cuanto a qué velocidad se produce la reacción. Consideremos una simple reacción



donde A es reactivo y B es el producto. La concentración de la sustancia reaccionante A decrece y la de B aumenta a medida que pasa el tiempo. La velocidad de reacción se define como la concentración cambiante de cualquiera reactivo o producto por unidad de tiempo. Por lo tanto,

$$\begin{aligned} \text{velocidad de reacción} &= \text{velocidad de desaparición de A} \\ &= \text{velocidad de aparición de B} \end{aligned}$$

o

$$\begin{aligned} \text{velocidad} &= -\frac{d[A]}{dt} \\ &= +\frac{d[B]}{dt}, \end{aligned}$$

donde [] representa la concentración. La tasa positiva significa que la concentración de B está aumentando con el tiempo, por ejemplo, un producto, y una tasa negativa significa que la concentración de A está disminuyendo con el tiempo, por ejemplo, un reactivo. La velocidad de reacción es influenciada por algunos factores tales como la temperatura, la presión, la concentración y catalizador, etc. Dependiendo de la relación entre la concentración y la velocidad de reacción, desarrollamos la ley y el orden de la velocidad de reacción.

2.3. Leyes de Velocidad y Orden de Reacción.

La relación exacta entre la concentración y la tasa se determina por la medición de la velocidad de reacción con diferente concentración inicial del reactivo. La relación entre la tasa y estas concentraciones a menudo se pueden expresar matemáticamente en la forma de un ecuación llamada ley de velocidad. Para numerosas reacciones, la velocidad de una reacción es directamente proporcional a la

concentración del reactivo, cada concentración se eleva a algunos de energía. Para una reacción



la velocidad de reacción con respecto a A o B se determina mediante la variación de la concentración de un reactivo, manteniendola de la otra constante. Por lo tanto la velocidad de reacción puede ser expresada como

$$\text{velocidad} = K[A]^k[B]^l, \quad (2.2)$$

donde k es constante, característica de una reacción particular, llamada la constante de velocidad o el coeficiente de tasa. La suma de la potencia de las concentraciones en la ley de velocidad, que es $(k + l)$ se conoce como el orden de la reacción.

2.4. Catálisis.

Berzelius (1836) se dió cuenta de que hay sustancias que aumentan la velocidad de una reacción sin que ellas mismas se consuman. Él creía que la función de dicha sustancia era para aflojar los lazos que mantiene los átomos en las moléculas que reaccionan entre sí. Así que llamó el término catálisis. No hay duda de que por lo general un catalizador acelera una reacción como se pensaba originalmente por Berzelius. Hoy en día un catalizador se define como una sustancia que altera la velocidad de una reacción química, en sí restante químicamente sin cambios al final de la reacción. Este proceso se llama catálisis. Por ejemplo, la conversión de sacarosa en dióxido de carbono (CO_2) y agua (H_2O) en presencia de oxígeno es un proceso altamente exergónico, la liberación de energía libre que se puede utilizar para pensar, moverse, el gusto y ver. Sin embargo, una bolsa de azúcar puede permanecer en el estante durante años sin ninguna conversión obvio para CO_2 y H_2O , este proceso químico es muy lento. Sin embargo, cuando la sacarosa es consumida por humanos, libera su energía química en cuestión de segundos. La diferencia es la catálisis. Sin catálisis, reacciones químicas, como la oxidación de sacarosa no podían producirse en un tiempo útil. Hay diferentes tipos de catálisis utilizados en la reacción. Un tipo de catálisis común se conoce como la catálisis enzimática, que es de gran interés biológico.

2.4.1. Catálisis Enzimática y Cinética.

Muchas de las reacciones bioquímicas de las que el metabolismo humano depende para avanzar a una velocidad útil son catalizadas por los compuestos orgánicos conocidos como enzimas. Las enzimas son las más notables, muy específicas en sus acciones y tienen un extraordinario poder catalítico. El estudio de las propiedades de estas enzimas puede conducir a una comprensión más profunda de la forma en que funciona el cuerpo en la salud y puede proporcionar datos importantes en cómo los problemas que surgen cuando los procesos habituales se ven afectados por la enfermedad, trauma o agentes ambientales. Si cualquier enzima actúa como un catalizador en una reacción al catalizador se le denomina catálisis enzimática. Por ejemplo, la reacción que produce la glucosa ($C_6H_{12}C_6$) y fructosa ($C_6H_{12}C_6$) de la sacarosa precursor ($C_{12}H_{22}C_{11}$), que requiere la presencia de la catálisis enzimática invertasa.



Ahora proporcionamos la cinética de las reacciones catalizadas por enzimas. Cinética de la enzima es el estudio de las reacciones químicas que son catalizadas por enzimas. En cinética de la enzima, la velocidad de reacción se mide y los efectos de variar las condiciones de la reacción investigada. El estudio de la cinética de una enzima de esta manera puede revelar el mecanismo catalítico de esta enzima, su papel en el metabolismo, cómo se controla su actividad, y cómo un fármaco o un veneno podrían inhibir la enzima. En 1913 un mecanismo general de la cinética enzimática fue desarrollado por el bioquímico estadounidense Leonor Michaelis y Maud Menten médico canadiense.

2.5. Mecanismo de Michaelis-Menten.

El mecanismo de Michaelis-Menten implica una enzima E primero combina reversiblemente con el sustrato S para formar el ES un complejo enzima-sustrato, a continuación, este último se descompone en una segunda etapa a partir de la enzima libre y el producto P . Este mecanismo está representado por la figura (2.1).

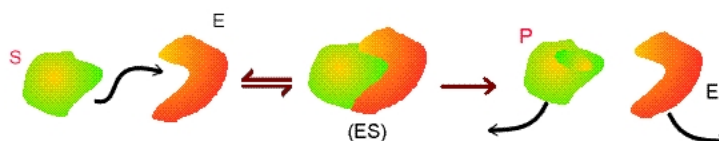


Figura 2.1: Mecanismo del Modelo de Michaelis-Menten.

Matemáticamente, este mecanismo puede escribirse como



donde k_1 , k_{-1} y k_2 denotan las constantes de velocidad de las tres reacciones. Aquí, la tasa de reacción inicial v se determina por la descomposición de ES complejo para dar el producto. La velocidad de reacción v es igual al producto de la concentración de la $[ES]$ complejo y k_2 es la constante de velocidad de descomposición complejo ES

$$v = k_2[ES]. \quad (2.4)$$

Porque $[ES]$ en la ecuación (2.4) no es fácil de medir experimentalmente, tenemos que empezar por encontrar una expresión alternativa para este término. En primer lugar se introduce el término $[Et]$, que representa la concentración total de enzima. Enzima libre se puede representar entonces por $[Et] - [ES]$. También, debido a que $[S]$ es normalmente mucho mayor que $[Et]$, la cantidad de sustrato obligado por la enzima en un momento dado es insignificante en comparación con el total de $[S]$. Con estas condiciones en mente, la expresión para v en términos de parámetros fácilmente medibles se describe de la siguiente manera: la velocidad de formación y la descomposición de ES se determina por la constante de velocidad k_1 (formación) y $k_{-1} + k_2$ (desglose de reactivos y productos, respectivamente), de acuerdo con las expresiones

$$\text{velocidad de formación de } ES = k_1([Et] - [ES])[S], \quad (2.5)$$

$$\text{índice de depresión de } ES = k_{-1}[ES] + k_2[ES]. \quad (2.6)$$

Ahora hacemos un supuesto importante, que la velocidad inicial de reacción refleja un estado estacionario en el que $[ES]$ es constante, la tasa de formación de ES es igual a la tasa de su degradación. Esto se llama suposición de estado estacionario (Lehninger et al., 2005). Las expresiones en las ecua-

ciones (2.5) y (2.6) se puede equiparar el estado estacionario,

$$\begin{aligned}
k_1 ([E_t] - [ES]) [S] &= k_{-1} [ES] + k_2 [ES] \\
\Rightarrow k_1 [E_t] [S] - k_1 [ES] [S] &= (k_{-1} + k_2) [ES] \\
\Rightarrow k_1 [E_t] [S] &= k_1 [ES] [S] + (k_{-1} + k_2) [ES] \\
\Rightarrow k_1 [E_t] [S] &= (k_1 [S] + k_{-1} + k_2) [ES] \\
\Rightarrow [ES] &= \frac{k_1 [E_t] [S]}{k_1 [S] + k_{-1} + k_2} \\
&= \frac{[E_t] [S]}{[S] + (k_{-1} + k_2)/k_1}.
\end{aligned} \tag{2.7}$$

La ecuación (2.7) se puede simplificar mediante la definición de una nueva constante

$$K_M = \frac{k_{-1} + k_2}{k_1}, \tag{2.8}$$

que es conocida como constante de Michaelis, y entonces la ecuación de estado estacionario (2.7) se convierte

$$[ES] = \frac{[E_t] [S]}{[S] + K_M}. \tag{2.9}$$

Ahora podemos expresar v en términos de $[ES]$. Sustituyendo el valor de $[ES]$ en la ecuación (2.4) da

$$v = \frac{k_2 [E_t] [S]}{[S] + K_M}. \tag{2.10}$$

La tasa aumenta a medida que la concentración de sustrato $[S]$ se incrementa (de acuerdo a la ley de velocidad) hasta un punto donde se dice que la enzima esta "saturado" con el sustrato. El tipo de medida v inicial llega a un valor máximo V_{max} y no se ve afectada por nuevos aumentos en la concentración de sustrato, debido a que el sustrato que está presente es más grande que la enzima. Por ejemplo, si una enzima con un peso molecular de 100,000 actúa sobre un sustrato con una peso molecular de 100 y ambos están presentes a una concentración de 1mg/ml, hay 1,000 mol de sustrato por cada mol de enzima (Murray et al., 2000). Es decir, la tasa máxima V_{max} se alcanza cuando los sitios de enzimas están saturados con sustrato, que es cuando $[S]$ es mucho mayor de K_M , por lo que $[S]/(K_M + [S])$ acercamiento a 1. Así

$$V_{max} = k_2 [E_t]. \tag{2.11}$$

Sustituyendo en la ecuación (2.10), obtenemos

$$v = \frac{V_{max} [S]}{[S] + K_M}. \quad (2.12)$$

Esta es la ecuación de Michaelis-Menten, la ecuación de velocidad para una reacción catalizada por la enzima de un sustrato. Se trata de una declaración de la relación cuantitativa entre v , V_{max} y $[S]$, todos relacionados a través de la K_M constante de Michaelis-Menten. La Figura (2.2) muestra la representación gráfica de la ecuación de Michaelis-Menten (2.12).

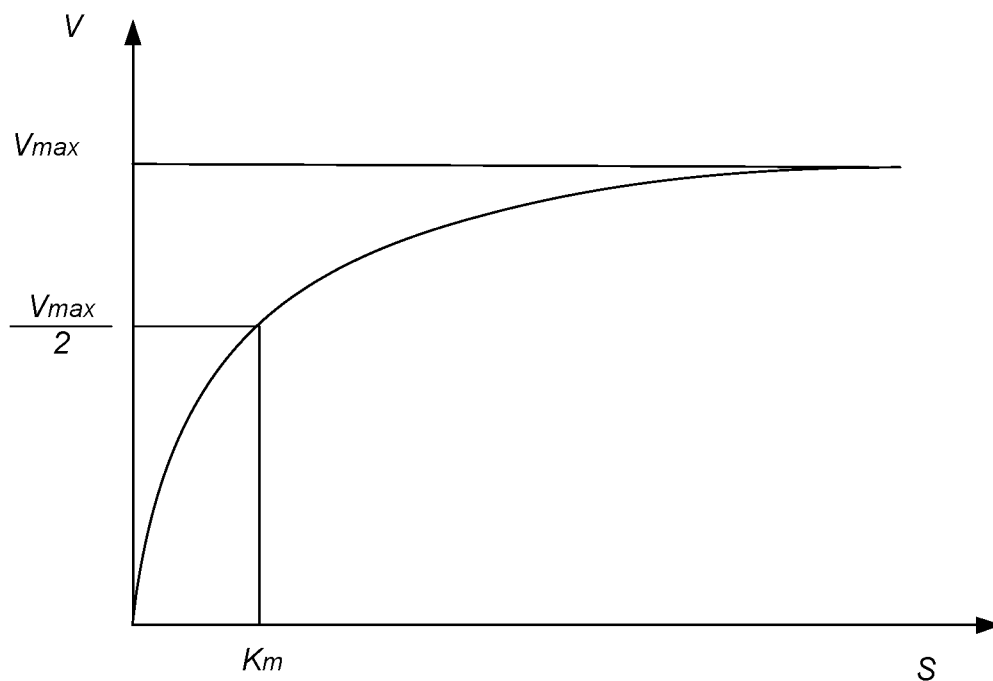


Figura 2.2: Efecto de la concentración del sustrato en la velocidad de una reacción catalizada por la enzima.

La Figura (2.2) muestra que los parámetros cinéticos definen los límites de la curva a alta y baja $[S]$.

1. A baja $[S]$, donde $[S] \ll K_M$, el término $[S]$ en el denominador de la ecuación (2.12) se convierte en insignificante, por lo que el término $[S]$ puede ser dado de baja del denominador.

A continuación, la ecuación (2.12) se simplifica a

$$\begin{aligned} v &= \frac{V_{max} [S]}{[S] + K_M} \\ \Rightarrow v &\approx \frac{V_{max} [S]}{K_M} \\ &= \frac{V_{max} [S]}{K_M}. \end{aligned} \tag{2.13}$$

y v exhibe una dependencia lineal en $[S]$.

2. A altas $[S]$, donde $[S] \gg K_M$, el término K_M en el denominador de la ecuación (2.12) se convierte en significativa, por lo que la expresión K_M puede ser dado de baja del denominador y la ecuación (2.12) se simplifica a

$$\begin{aligned} v &= \frac{V_{max} [S]}{[S] + K_M} \\ \Rightarrow v &\approx \frac{V_{max} [S]}{[S]} \\ &= V_{max}. \end{aligned} \tag{2.14}$$

y v es consistente con alto $[S]$.

La ecuación de Michaelis-Menten por lo tanto, es consistente con la dependencia observada de v en $[S]$, y la forma de la curva se define por los términos V_{max}/K_M a baja $[S]$ y V_{max} en alto $[S]$. Esa es una definición muy útil y práctica de la K_M . K_M es equivalente a la concentración de sustrato en la que v es un medio V_{max} .

2.6. Estimación de los Parámetros del Modelo de Michaelis-Menten.

Con el fin de estimar los parámetros de K_M y V_{max} , los investigadores observan las velocidades de las reacciones, v_i en una secuencia de concentraciones de sustrato de x_i , $i = 1, \dots, N$ y luego la ecuación (2.12) se ajusta a estos datos. El investigador puede elegir las concentraciones de sustrato, principalmente mediante la aplicación de la teoría de diseño óptimo al problema. Con el fin de hacer esto es necesario para ser más precisos acerca de cómo se monta la ecuación (2.12) para los datos. En el primer caso, se supone que

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i, \quad (2.15)$$

donde $\varepsilon_i \sim N(0, \sigma^2)$; y sea

$$\eta(x_i, \theta) = \frac{\theta_1 x_i}{\theta_2 + x_i}, \quad (2.16)$$

donde $\theta = (\theta_1, \theta_2)$.

La ecuación (2.15) es el modelo de regresión no lineal, porque diferenciando (2.16) con respecto a θ_1 y θ_2 da

$$\eta(x_i, \theta_1) = \frac{x_i}{\theta_2 + x_i}, \quad (2.17)$$

$$\eta(x_i, \theta) = -\frac{\theta_1 x_i}{(\theta_2 + x_i)^2}, \quad (2.18)$$

y ambas derivadas implican al menos uno de los parámetros.

Capítulo 3

Modelos Estadísticos.

Diseños experimentales óptimos dependen del modelo, aunque por lo general no para los modelos lineales, en los valores de los parámetros de los modelos. Esta sección pretende dar una discusión de la elección de una forma apropiada para un modelo.

3.1. Modelos Lineales.

En el modelo estadístico, la variable de respuesta es por lo general denota por y , mientras que las variables explicativas k son típicamente representados por x_1, \dots, x_k . La dependencia de la variable respuesta de los niveles de los factores explicativos se modela mediante la función de respuesta

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \\ &= f^T(x) \beta, \end{aligned} \tag{3.1}$$

donde f es la expansión polinómica de las variables explicativas $x = (x_1, \dots, x_k)$ y β es el vector de parámetros de $p \times 1$. Debido a que la observación experimental está sujeta a variación aleatoria, el modelo estadístico añade un término de error aleatorio ε para la función de respuesta. La i -ésima observación experimental a continuación, se puede escribir como

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \\ &= f^T(x_i) \beta + \varepsilon_i, \quad i = 1, \dots, n, \end{aligned} \tag{3.2}$$

con $\varepsilon_i \sim N(0, \sigma^2)$. En esta expresión, x_i representa el ajuste de las variables explicativas en la i -ésima prueba experimental y se conoce como el punto de diseño o el tratamiento correspondiente a la i -ésima observación.

3.1.1. Estimación del Modelo Lineal.

Se puede aplicar el método de mínimos cuadrados para estimar los coeficientes del modelo lineal (3.1). Supongamos que se dispone de $n > k$ observaciones, y sea y_i la i -ésima respuesta observada, y x_{ij} la i -ésima observación o nivel del regresor x_j . Los datos aparecerán como en la tabla (3.1). Se supone que el término del error ε del modelo tiene $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2$, y que los errores no están correlacionados.

Observación	Respuesta	Regresores				
i	y	x_1	x_2	...	x_k	
1	y_1	x_{11}	x_{21}	...	x_{1k}	
2	y_2	x_{21}	x_{22}	...	x_{2k}	
\vdots	\vdots	\vdots	\vdots		\vdots	
n	y_n	x_{n1}	x_{n2}	...	x_{nk}	

Tabla 3.1: Datos para el modelo lineal.

En esta sección se supondrá que las variables regresoras x_1, x_2, \dots, x_k son fijas, es decir, que son matemáticas o no aleatorias, y que se miden sin error. Esto es realmente importante, porque cuando se toman datos del modelo lineal en un estudio observacional, algunos o la mayor parte de los regresores son variables aleatorias. Cuando los datos son el resultado de un experimento diseñado es más probable que las x sean variables fijas. Cuando las x son variables aleatorias sólo es necesario que las observaciones con cada regresor sean independientes, y que la distribución no dependa de los coeficientes del modelo lineal (las β) o de σ^2 . Cuando se prueban hipótesis o se establecen intervalos de confianza, se debe suponer que la distribución condicional de y dadas x_1, x_2, \dots, x_k es normal con promedio $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ y varianza σ^2 .

La función de mínimos cuadrados es

$$\begin{aligned}
 S(\beta) &= \sum_{i=1}^n \varepsilon_i^2 \\
 &= \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2.
 \end{aligned} \tag{3.3}$$

Se debe minimizar la función S respecto a $\beta_0, \beta_1, \dots, \beta_k$. Los estimadores de $\beta_0, \beta_1, \dots, \beta_k$ por mínimos

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \quad (3.7)$$

En general, y es un vector de $n \times 1$ de las observaciones, X es una matriz de $n \times p$ de los niveles de las variables regresoras, β es un vector de $p \times 1$ de los coeficientes del modelo lineal y ε es un vector de $n \times 1$ de errores aleatorios.

Se desea determinar el vector $\hat{\beta}$ de estimadores de mínimos cuadrados que minimice

$$\begin{aligned} S(\beta) &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \varepsilon^T \varepsilon \\ &= (y - X\beta)^T (y - X\beta) \end{aligned} \quad (3.8)$$

Nótese que $S(\beta)$ se puede expresar como sigue:

$$\begin{aligned} S(\beta) &= y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta \\ &= y^T y - 2\beta^T X^T y + \beta^T X^T X\beta \end{aligned} \quad (3.9)$$

ya que $\beta^T X^T y$ es una matriz de 1×1 , es decir, un escalar, y que su transpuesta $(\beta^T X^T y)^T = y^T X\beta$ es el mismo escalar. Los estimadores de mínimos cuadrados deben satisfacer

$$\left. \frac{\partial L}{\partial \beta} \right|_{\hat{\beta}} = -X^T y + 2X^T X\beta = 0, \quad (3.10)$$

lo que simplifica a

$$X^T X\beta = X^T y. \quad (3.11)$$

Las ecuaciones (3.11) son las ecuaciones normales de mínimos cuadrados. Son la forma matricial de la presentación escalar, ecuaciones (3.6). Para resolver las ecuaciones normales se multiplican ambos lados de (3.11) por la inversa de $X^T X$. Así, el estimador de β por mínimos cuadrados es

$$\hat{\beta} = (X^T X)^{-1} X^T y. \quad (3.12)$$

siempre y cuando exista la matriz inversa $(X^T X)^{-1}$. La matriz $(X^T X)^{-1}$ siempre existe si los regresores son linealmente independientes, esto es, si ninguna columna de la matriz X es una combinación lineal de las demás columnas.

La matriz de varianza y covarianza se puede expresar como

$$\text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}, \quad (3.13)$$

y la matriz de información sobre el parámetro β viene dado por

$$M = \sigma^{-2} (X^T X). \quad (3.14)$$

Una vez que el modelo es estimado, se utiliza para predecir la respuesta de varias combinaciones x de los factores experimentales. La respuesta predicha está dada por

$$y(x) = f^T(x) \beta, \quad (3.15)$$

y la varianza puede ser escrita como

$$\text{Var}[y(x)] = \sigma^2 f^T(x) (X^T X)^{-1} f(x). \quad (3.16)$$

3.2. Modelos No Lineales.

Hay muchas situaciones en las que puede no ser lo adecuado un modelo lineal de regresión, por ejemplo, cuando se tiene el conocimiento directo de la forma de la relación entre la variable de respuesta y los regresores, quizá con base en la teoría que gobierna los fenómenos. La relación real entre la respuesta y los regresores puede ser una ecuación diferencial, o su solución; con frecuencia, esa ecuación conducirá

a un modelo de forma no lineal. Cualquier modelo que no sea lineal en los parámetros desconocidos es un modelo de regresión no lineal. En general, se escribirá el modelo de regresión no lineal en la forma

$$y = \eta(x, \theta) + \varepsilon, \quad (3.17)$$

donde θ es el vector de parámetros de $p \times 1$ y $\varepsilon_i \sim N(0, \sigma^2)$. También, se supondrá habitualmente que los errores tienen distribución normal, como en la regresión lineal. Como

$$\begin{aligned} E(y) &= E[\eta(x, \theta) + \varepsilon] \\ &= \eta(x, \theta) \end{aligned} \quad (3.18)$$

a la función $\eta(x, \theta)$ se le llama función de valor esperado para el modelo no lineal. Esto se parece mucho al caso del modelo lineal, excepto que ahora la función de valor esperado es una función no lineal de los parámetros.

En un modelo de regresión no lineal, al menos una de las derivadas de la función de valor esperado con respecto a los parámetros depende de cuando menos uno de los parámetros. En la regresión lineal, esas derivadas no son funciones de los parámetros desconocidos. Para ilustrar lo anterior, consideremos el modelo lineal (3.1), cuya función de valor esperado $\eta(x, \theta) = \beta_0 + \sum_{j=1}^k \beta_j x_j$. Ahora bien,

$$\frac{\partial \eta(x, \theta)}{\partial \beta_j} = x_j, \quad j = 0, 1, \dots, k \quad (3.19)$$

siendo $x_0 \equiv 1$. Obsérvese que en el caso lineal las derivadas no son funciones de las β . Ahora considérese el modelo no lineal de Michaelis-Menten de cinética química, para relacionar la velocidad inicial de una reacción enzimática con la concentración x del sustrato. El modelo es

$$\begin{aligned} y &= \eta(x, \theta) + \varepsilon \\ &= \frac{\theta_1 x}{\theta_2 + x} + \varepsilon \end{aligned} \quad (3.20)$$

Las derivadas de la función de valor esperado con respecto a θ_1 y θ_2 son

$$\frac{\partial \eta(x, \theta)}{\partial \theta_1} = \frac{x}{\theta_2 + x}, \quad (3.21)$$

y

$$\frac{\partial \eta(x, \theta)}{\partial \theta_2} = -\frac{\theta_1 x}{(\theta_2 + x)^2}, \quad (3.22)$$

Como las derivadas son función de los parámetros desconocidos θ_1 y θ_2 , el modelo es no lineal .

3.2.1. Estimación del Modelo No lineal.

Supóngase que una muestra de n observaciones de la respuesta y los regresores es $y_i, x_{i1}, x_{i2}, \dots, x_{ik}$ para $i = 1, 2, \dots, n$. El estimador de mínimos cuadrados de θ , denotado por $\hat{\theta}$ minimiza la suma de cuadrados del error,

$$S(\theta) = \sum_{i=1}^n [y_i - \eta(x_i, \theta)]^2. \quad (3.23)$$

Para encontrar el estimador de mínimos cuadrados $\hat{\theta}$ es necesario diferenciar la ecuación (3.23) con respecto a θ . Esto proporciona p ecuaciones normales, las que deben ser resueltas para $\hat{\theta}$. Las ecuaciones normales son tomadas de la forma

$$\sum_{i=1}^n [y_i - \eta(x_i, \theta)] \left[\frac{\partial \eta(x_i, \theta)}{\partial \theta} \right]_{\theta=\hat{\theta}} = 0. \quad (3.24)$$

para $j = 1, 2, \dots, p$. En un modelo de regresión no lineal, las derivadas entre corchetes grandes serán funciones de los parámetros desconocidos. Además, la función del valor esperado también es no lineal, por lo que las ecuaciones normales pueden ser muy difíciles de resolver.

Para el modelo no lineal de regresión, en la ecuación (3.20) las ecuaciones normales de mínimos cuadrados para este modelo son

$$\begin{aligned} \sum_{i=1}^n \left[y_i - \frac{\hat{\theta}_1 x}{\hat{\theta}_2 + x} \right] \left[\frac{x}{\hat{\theta}_2 + x} \right] &= 0 \\ \sum_{i=1}^n \left[y_i - \frac{\hat{\theta}_1 x}{\hat{\theta}_2 + x} \right] \left[-\frac{\hat{\theta}_1 x}{(\hat{\theta}_2 + x)^2} \right] &= 0 \end{aligned} \quad (3.25)$$

Estas ecuaciones son no lineales en $\hat{\theta}_1$ y $\hat{\theta}_2$, y no existe solución sencilla de forma cerrada. En general, se deben usar métodos iterativos para determinar los valores $\hat{\theta}_1$ y $\hat{\theta}_2$. Para complicar más el problema, a veces hay varias soluciones de las ecuaciones normales, lo que nos lleva a que hay diferentes valores estacionarios para la función suma de cuadrados de residuales, $S(\theta)$.

3.3. Transformación a un Modelo Lineal.

A veces es útil considerar una transformación que induzca la linealidad en la función de valor esperado del modelo. Por ejemplo, se tiene el modelo (3.20), donde se ve que la función se puede linealizar con facilidad, ya que

$$\begin{aligned}
 \frac{1}{\eta(x, \theta)} &= \frac{\theta_2 + x}{\theta_1 x} + \frac{1}{\varepsilon} \\
 &= \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} \frac{1}{x} + \frac{1}{\varepsilon} \\
 &= \beta_0 + \beta_1 u + \varepsilon^*.
 \end{aligned}
 \tag{3.26}$$

Consecuentemente es muy atractivo usar un modelo lineal para estimar β_0 y β_1 . Sin embargo, los estimados por mínimos cuadrados de los parámetros de la ecuación (3.26) no serán, en general, equivalentes a los estimados no lineales de parámetros en el modelo original, ecuación (3.20). La razón es que en el modelo no lineal original, los mínimos cuadrados implican la minimización de la suma de los residuales al cuadrado respecto a y , mientras que en el modelo transformado, ecuación (3.26), se está minimizando la suma de los residuales al cuadrado respecto a la inversa de y , si ε^* sigue una distribución normal, se aplicarán todas las propiedades del modelo lineal.

Un modelo no lineal que se puede transformar en una forma lineal equivalente se llama intrínsecamente lineal, sin embargo, lo importante suele implicar la estructura de los errores: ¿se aplican las suposiciones estándar de los errores al modelo no lineal original o alinealizado? Esta pregunta a veces no es fácil de contestar.

Los datos en la tabla (3.2) corresponden a la velocidad inicial de reacción para una enzima tratada con puromicina (Bates y Watts (1988)), y se grafican en la figura (3.1).

Concentración del sustrato (ppm)	Velocidad $\frac{\text{cantidad}}{\text{min}^2}$
0.02	47 76
0.06	97 107
0.11	123 139
0.22	152 159
0.56	191 201
1.10	200 207

Tabla 3.2: Velocidad de reacción y concentración de sustrato en el experimento de puromicina.

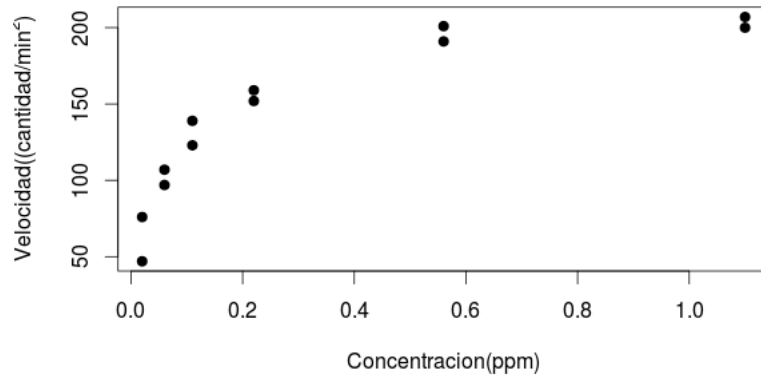


Figura 3.1: Gráfica de velocidad de reacción en función de concentración de sustrato, para el experimento de puromicina.

por lo que se puede tratar de ajustar el modelo lineal

$$y^* = \beta_0 + \beta_1 u + \varepsilon^*, \quad (3.27)$$

en donde $y^* = 1/y$ y $u = 1/x$. El ajuste resultante por mínimos cuadrados es

$$\hat{y}^* = 0.005107 + 0.0002472u. \quad (3.28)$$

La figura (3.2) muestra un diagrama de dispersión de los datos transformados y^* y u , con la recta superpuesta. Como hay réplicas en los datos, es fácil ver, en la figura (3.1), que la varianza de los datos originales es aproximadamente constante, mientras que según la figura (3.2), la suposición de varianza constante en la escala transformada es irrazonable.

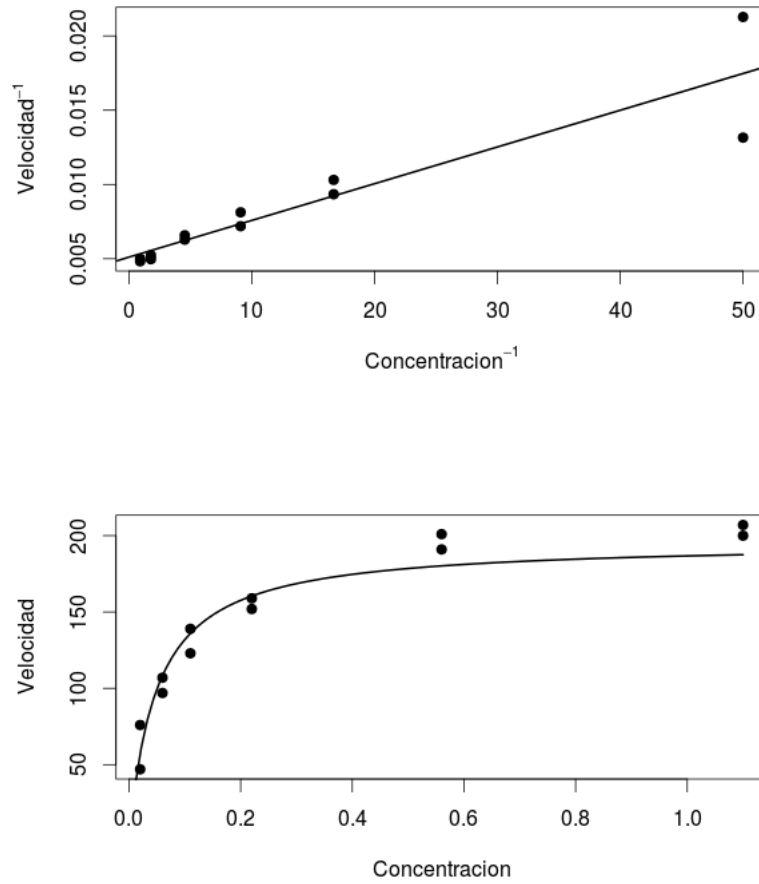


Figura 3.2: a) Gráfica del inverso de la velocidad en función del inverso de la concentración, para los datos de puromicina. b) Curva ajustada en la escala original.

Ahora bien, como

$$\beta_0 = \frac{1}{\theta_1} \quad \text{y} \quad \beta_1 = \frac{\theta_2}{\theta_1},$$

entonces

$$0.005107 = \frac{1}{\hat{\theta}_1} \quad \text{y} \quad 0.0002472 = \frac{\hat{\theta}_2}{\hat{\theta}_1},$$

por lo que se pueden estimar θ_1 y θ_2 en el modelo original como sigue:

$$\hat{\theta}_1 = 195.81 \quad \text{y} \quad \hat{\theta}_2 = 0.04841.$$

La figura (3.2) muestra la curva ajustada en la escala original, junto con los datos. Obsérvese en

la figura que la asíntota ajustada es demasiado baja. La varianza en los puntos replicados ha sido distorsionada por la transformación, por lo que las corridas con baja concentración (alta concentración recíproca) dominan el ajuste por mínimos cuadrados, y en consecuencia el modelo no se ajusta bien a los datos cuando las concentraciones son altas.

3.4. Estimación de los Parámetros en un Sistema No Lineal.

Un método muy usado en los algoritmos de cómputo para regresión no lineal es la linealización de la función no lineal, seguida por el método iterativo de Gauss-Newton para estimar parámetros. La linealización se hace con un desarrollo en serie de Taylor de $f(x_i, \theta)$ respecto al punto $\theta_0^T = [\theta_{10}, \theta_{20}, \dots, \theta_{p0}]$, reteniendo sólo los términos lineales. Así se obtiene

$$y_i = \eta(x_i, \theta_0) + \left. \frac{\partial \eta(x_i, \theta)}{\partial \theta_j} \right|_{\theta=\theta_0} (\theta_j - \theta_{j0}) \quad (3.29)$$

Si se definen

$$\begin{aligned} \eta_i^0 &= \eta(x_i, \theta_0) \\ \beta_j^0 &= \theta_j - \theta_{j0} \\ Z_{ij}^0 &= \left[\frac{\partial \eta(x_i, \theta)}{\partial \theta_j} \right]_{\theta=\theta_0} \end{aligned} \quad (3.30)$$

se nota que el modelo de regresión no lineal se puede escribir en la forma

$$y_i - \eta_i^0 = \sum_{j=1}^p \beta_j^0 Z_{ij}^0 + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (3.31)$$

Esto es, ahora se cuenta con un modelo lineal. Se acostumbra llamar θ_0 a los valores iniciales de los parámetros. La ecuación (3.31) se puede escribir en la forma

$$y_0 = Z_0 \beta_0 + \varepsilon, \quad (3.32)$$

por lo que el estimado de β_0 es

$$\begin{aligned}\hat{\beta}_0 &= (Z_0^T Z_0)^{-1} Z_0^T y_0 \\ &= (Z_0^T Z_0)^{-1} Z_0^T (y - \eta_0),\end{aligned}\tag{3.33}$$

Ahora bien, como $\beta_0 = \theta - \theta_0$, se puede definir

$$\hat{\theta}_1 = \hat{\beta}_0 + \theta_0,\tag{3.34}$$

como los estimados corregidos de θ . A veces a $\hat{\beta}_0$ se le llama vector de incrementos. Ahora se pueden reemplazar los estimados corregidos θ_1 en la ecuación (3.29) (en los mismos papeles que desempeñaban los estimados iniciales θ_0), para a continuación producir otro conjunto de estimados corregidos, por ejemplo, $\hat{\theta}_2$, y así sucesivamente.

En general, en la k -ésima iteración se tiene

$$\begin{aligned}\hat{\theta}_{k+1} &= \hat{\theta}_k + \hat{\beta}_k \\ &= \hat{\theta}_k + (Z_k^T Z_k)^{-1} Z_k^T (y - \eta_k),\end{aligned}\tag{3.35}$$

siendo

$$\begin{aligned}Z_k &= [Z_{ij}^k] \\ \eta_k &= [\eta_1^k, \eta_2^k, \dots, \eta_n^k] \\ \hat{\theta}_k &= [\theta_{1k}, \theta_{2k}, \dots, \theta_{pk}]^T.\end{aligned}\tag{3.36}$$

Este procedimiento de iteración continúa hasta la convergencia, esto es, hasta que

$$\left| \frac{\hat{\theta}_{j,k+1} - \hat{\theta}_{jk}}{\hat{\theta}_{jk}} \right| < \delta, \quad j = 1, 2, \dots, p\tag{3.37}$$

en donde δ es un número pequeño, por ejemplo 1.0×10^{-6} . En cada iteración se debe evaluar la suma de cuadrados de residuales, $S(\hat{\theta}_k)$ para asegurar que se haya obtenido una reducción en su valor.

Bates y Watts (1988) aplican el método de Gauss-Newton para ajustar el modelo de Michaelis-Menten a los datos de la puromicina, en la tabla (3.2), con los valores iniciales $\theta_{10} = 205$ y $\theta_{20} = 0.08$. En este

punto de partida, la suma de cuadrados de residuales es $S(\theta_0) = 3155$; los datos, los valores ajustados, los residuales y las derivadas, evaluadas en cada observación, se ven en la tabla (3.3). Para ilustrar cómo se calcularon las cantidades necesarias, obsérvese que

$$\frac{\partial \eta(x, \theta)}{\partial \theta_1} = \frac{x}{\theta_2 + x},$$

y que

$$\frac{\partial \eta(x, \theta)}{\partial \theta_2} = -\frac{\theta_1 x}{(\theta_2 + x)^2},$$

y como la primera observación de x es $x_1 = 0.02$, entonces

$$\begin{aligned} \eta_1^0 &= \left. \frac{\theta_1 x}{\theta_2 + x} \right|_{\theta_1=205, \theta_2=0.08} \\ &= \frac{(205)(0.02)}{0.08 + 0.02} \\ &= 41.00 \end{aligned}$$

$$\begin{aligned} Z_{11}^0 &= \left. \frac{x}{\theta_2 + x} \right|_{\theta_2=0.08} \\ &= \frac{0.02}{0.08 + 0.02} \\ &= 0.2 \end{aligned}$$

$$\begin{aligned} Z_{12}^0 &= \left. -\frac{\theta_1 x}{(\theta_2 + x)^2} \right|_{\theta_1=205, \theta_2=0.08} \\ &= -\frac{(205)(0.02)}{(0.08 + 0.02)^2} \\ &= -410.00 \end{aligned}$$

Ahora se agrupan las derivadas Z_{ij}^0 en la matriz Z_0 y el vector de incrementos se calcula con la ecuación 3.33, resultando

$$\beta_0 = \begin{bmatrix} 8.03 \\ -0.017 \end{bmatrix}, \quad (3.38)$$

i	x_i	y_i	η_i^0	$y_i - \eta_i^0$	Z_{i1}^0	Z_{i2}^0
1	0.02	47	41.00	6.00	0.2000	-410.00
2	0.02	76	41.00	35.00	0.2000	-410.00
3	0.06	97	87.86	9.14	0.4286	-627.55
4	0.06	107	87.86	19.14	0.4286	-627.55
5	0.11	123	118.68	4.32	0.5789	-624.65
6	0.11	139	118.68	20.32	0.5789	-624.65
7	0.22	152	150.33	8.67	0.7333	-501.11
8	0.22	159	150.33	1.67	0.7333	-501.11
9	0.56	191	179.38	11.62	0.8750	-280.27
10	0.56	201	179.38	21.62	0.8750	-280.27
11	1.10	200	191.10	15.90	0.9322	-161.95
12	1.10	207	191.10	8.90	0.9322	-161.95

Tabla 3.3: Datos, valores ajustados y derivadas para el experimento de puromicina en $\hat{\theta}_0^T = (205, 0.08)^T$.

El estimado corregido de $\hat{\theta}_1$ de la ecuación (3.34) es

$$\begin{aligned}
\hat{\theta}_1 &= \hat{\beta}_0 + \theta_0 \\
&= \begin{bmatrix} 8.03 \\ -0.017 \end{bmatrix} + \begin{bmatrix} 205 \\ 0.08 \end{bmatrix} \\
&= \begin{bmatrix} 213.03 \\ 0.063 \end{bmatrix}, \tag{3.39}
\end{aligned}$$

La suma de cuadrados de residuales, en este punto, es $S(\hat{\theta}_1) = 1206$, bastante menor que $S(\theta_0)$. Por consiguiente, se adopta $\hat{\theta}_1$ como el estimado corregido de θ y se corre otra iteración. El algoritmo de Gauss-Newton convergió hacia $\hat{\theta}^T = (212.7, 0.0641)^T$, con $S(\hat{\theta}) = 1195$.

La figura (3.3) muestra el modelo ajustado. Obsérvese que este modelo no lineal produce un ajuste mucho mejor a los datos que la transformación seguida por el modelo lineal (compárense las Figs. (3.3) y (3.2)).

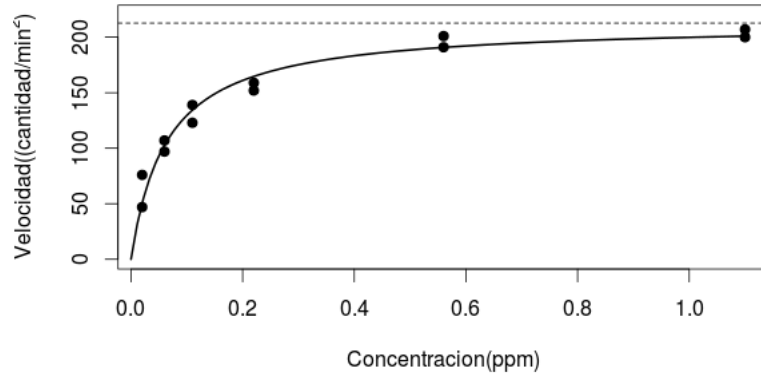


Figura 3.3: Gráfica del modelo ajustado del modelo no lineal de Michaelis-Menten.

La matriz de varianza y covarianza se puede expresar como

$$\text{Var}(\theta) = \sigma^{-2} (Z_0^T Z_0)^{-1}, \quad (3.40)$$

y la matriz de información sobre el parámetro β desconocido viene dado por

$$M = \sigma^{-2} (Z_0^T Z_0). \quad (3.41)$$

Capítulo 4

Criterios de Optimalidad.

4.1. Introducción.

¿Qué entendemos por diseño óptimo de un experimento? ¿Cómo entender la expresión “el mejor de los diseños posibles”? Estas preguntas no tienen una respuesta fácil. Si bien es verdad que nos interesará el diseño que haga mínima la varianza, también es cierto que un diseño puede hacer mínima la varianza para un funcional lineal, y excesivamente grande para otro.

El diseño óptimo, también conocido como diseños óptimos, es una clase de diseño experimental que se optimiza con respecto a algún criterio estadístico. Un diseño no óptimo requiere un mayor número de corridas experimentales para estimar los parámetros con la misma precisión que un diseño óptimo, es decir, un diseño óptimo puede reducir los costos de la experimentación. La optimalidad de un diseño depende del modelo estadístico que se va a utilizar se analiza y se evalúa con respecto a un criterio estadístico. Especificar un modelo apropiado y especificar una función criterio adecuada requiere la comprensión de la teoría estadística y el conocimiento práctico con el diseño de experimentos.

4.2. Teoría del Diseño Óptimo

El propósito de la teoría de diseño óptimo es determinar los valores de las variables explicativas x cada vez que una observación experimental se hizo (Goos, 2002). En la literatura, el problema matemático de encontrar el diseño óptimo se viró en dos formas conocidas como el diseño continuo y exacto.

4.2.1. Diseños Continuos y Diseños Exactos.

En el diseño continuo el número de la observación en un determinado punto de diseño debe ser un entero. En el diseño continuo están representados por una medida ξ sobre el diseño en la región \mathcal{X} . Si el diseño tiene observaciones en k puntos distintos, se denota por

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_k \\ w_1 & w_2 & \dots & w_k \end{array} \right\}, \quad (4.1)$$

donde x_1, x_2, \dots, x_k son los niveles del factor experimental en cada punto de diseño y w_1, w_2, \dots, w_k son los pesos asociados con cada punto de diseño. Así ξ es una medida, $\int_{\mathcal{X}} \xi dx = 1$ y $w_i \geq 0$ para toda i . Un diseño exacto con n observaciones puede ser denotado como

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_k \\ n_1 & n_2 & \dots & n_k \end{array} \right\}, \quad (4.2)$$

donde n_1, n_2, \dots, n_k es el número entero de ensayos en x_1, x_2, \dots, x_k y $\sum_{i=1}^k n_i = n$.

Los diseños continuos son de gran importancia práctica en los casos en que ξ resulta dar pesos racionales simples para un pequeño número de puntos de diseño, ya que esto permite a los diseñadores elegir, para algún n , que son conocidos por ser óptimo. Para los modelos no lineales los pesos son números racionales raramente simples. Para n grande por lo general es posible encontrar ξ_n que es una buena aproximación a ξ , en cuyo caso ξ_n da un diseño exacto casi óptimo. Para las pequeñas o moderadas n puede que no sea posible encontrar ξ_n que de buena aproximación a ξ . Se ha demostrado, por ejemplo, por Bohachevsky et al. (1986), que aproximar ξ lo mejor posible no es suficiente, ya que los diseños obtenidos pueden ser lejos de ser óptimos. Por lo tanto, para la experimentación práctica se necesitan diseños exactos y en este informe se utilizan diseños óptimos exactos.

4.2.2. Matriz de Información para los Diseños Continuos y los Diseños Exactos.

Para un diseño continuo ξ la matriz de información es definida como

$$\begin{aligned}
 M(\xi) &= \int_{\mathcal{X}} f(x) f^T(x) \xi(dx) \\
 &= \sum_{i=1}^k w_i M(\bar{\xi}_i) \\
 &= \sum_{i=1}^k w_i f(x_i) f^T(x_i),
 \end{aligned} \tag{4.3}$$

donde la medida $\bar{\xi}_i$ pone unidades de masa a cada x_i . Debido a la presencia de los pesos w_i , la última forma de (4.3), resumido en los puntos de diseño k , se convierte en la matriz de información del diseño exacto ξ_n . Es decir

$$M(\xi_n) = \sigma^{-2} \sum_{i=1}^k n_i f(x_i) f^T(x_i). \tag{4.4}$$

La varianza de la respuesta predicha para un diseño de n ensayos se da en (3.13) como

$$\text{var}[\hat{y}(x)] = \sigma^2 f^T(x) (X^T X)^{-1} f(x). \tag{4.5}$$

Para diseños continuos la varianza normalizada de la respuesta predicha es

$$\begin{aligned}
 d(x, \xi) &= \frac{\text{var}[\hat{y}(x)]}{\sigma^2} \\
 &= \frac{\sigma^2 f^T(x) M(\xi)^{-1} f(x)}{\sigma^2} \\
 &= f^T(x) M(\xi)^{-1} f(x).
 \end{aligned} \tag{4.6}$$

una función tanto del diseño ξ y el punto en el que se hizo la predicción. Si el diseño es exacto, la varianza normalizada de la respuesta predicha es

$$\begin{aligned}
 d(x, \xi_n) &= n\sigma^{-2} f^T(x) M(\xi_n)^{-1} f(x) \\
 &= n\sigma^{-2} \text{var}[\hat{y}(x)].
 \end{aligned} \tag{4.7}$$

4.3. Criterios de Optimalidad.

Necesitamos por tanto elegir un criterio que nos sirva para buscar el mejor diseño en algún sentido. Su elección dependerá de los intereses que se busquen al realizar el experimento, de la facilidad de cálculo, o de otros aspectos más o menos subjetivos.

Damos ahora una primera definición de lo que va a ser una función criterio. Más tarde daremos una definición más completa. Para diferenciar una de otra, a la segunda la llamaremos función criterio convexa.

Definición 1. Diremos que una función:

$$\Phi : \mathcal{M} \rightarrow \mathbb{R} \cup \{+\infty\}$$

acotada inferiormente es una función criterio si se cumple lo siguiente:

$$var_{\xi}g \leq var_{\eta}g \text{ para todo funcional } g \Rightarrow \Phi[M(\xi)] \leq \Phi[M(\eta)]$$

Diremos entonces que se trata de un criterio de Φ -optimización. Un diseño que minimice $\Phi[M(\xi)]$ se denominará diseño Φ -óptimo.

Proposición 2. *Toda función criterio Φ , tal y como se ha definido anteriormente es siempre decreciente, en el sentido siguiente:*

$$M(\xi) \geq M(\eta) \Rightarrow \Phi[M(\xi)] \leq \Phi[M(\eta)]$$

Observación 3. Puede ocurrir que dos funciones criterio den lugar a un mismo criterio de optimización. Es precisamente lo que demostrará el teorema de equivalencia para algunos criterios que veremos más adelante.

Definición 4. Llamaremos $\mathcal{M}_+ = \{M \in \mathcal{M} : \det M > 0\}$. $\mathcal{L}(\mathcal{M})$ es un subespacio vectorial de las matrices simétricas de orden m generado por el conjunto \mathcal{M} . Si Φ es una función criterio definimos los conjuntos

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

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

Proposición 5. Si Φ es una función convexa entonces el conjunto Ξ^* es convexo.

Demostración. Supongamos que:

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

entonces, por ser Φ convexa tendremos

$$\begin{aligned}\Phi \{ \Phi [M ((1 - \beta) \xi_1 + \beta \xi_2)] \} &= \Phi \{ (1 - \beta) M (\xi_1) + \beta M (\xi_2) \} \\ &\leq (1 - \beta) \Phi [M (\xi_1)] + \beta \Phi [M (\xi_2)] \\ &= \min_{\xi \in \Xi} \Phi [M (\xi)]\end{aligned}$$

□

Observación 6. En algunas ocasiones no es necesario estimar más que un subconjunto de los parámetros o algunas funciones de ellos, de modo que la función criterio restringe su atención a dichos parámetros. Hablaremos entonces de criterios de optimización parcial o criterios de optimización singulares. Un ejemplo de criterio de optimización parcial viene dado por la función:

$$\Phi : M (\xi) \rightarrow var_{\xi} g$$

Aunque la función criterio global sea continua, la función criterio parcial correspondiente no siempre es continua en todo \mathcal{M} . Además el diseño Φ -óptimo tiene matriz de información singular en algunos casos. Si los parámetros en los que estamos interesados son los s primeros entonces la función criterio centrará su atención en las varianzas y covarianzas de estos s parámetros, es decir, en la caja superior izquierda de orden $s \times s$ de la matriz de información.

A partir de los criterios que se estudiarán en el siguiente capítulo pueden definirse criterios parciales de manera natural.

4.4. Funciones criterios y sus propiedades.

Para la definición y propiedades de algunas funciones criterio necesitaremos de las siguientes definiciones:

Definición 7. Sea $M_m(\mathbb{R}) \cong \mathbb{R}^{m \times m}$ el conjunto de las matrices cuadradas con términos reales de orden m . En este espacio vectorial se puede definir el siguiente producto escalar:

$$\langle A, B \rangle = \text{tr} A^T B$$

de donde la norma de una matriz cuadrada cualquiera de orden m vendría dado por:

$$\|A\|^2 = \text{tr} A^T A$$

Sabemos que cualquier otra norma definida en este espacio euclídeo es topológicamente equivalente a ésta. En lo sucesivo utilizaremos esta norma.

Definición 8. Sea Φ una función definida en un entorno a la matriz A en el espacio $\mathcal{M}_m(\mathbb{R})$. Se define el gradiente de Φ en la matriz A como la matriz de componentes:

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

Damos ya la definición de las funciones criterio más utilizadas:

4.4.1. D-optimalidad.

Definición 9. El criterio de D-optimización viene definido por la función criterio siguiente:

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

Proposición 10. Φ_D tiene las siguientes propiedades:

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

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

Observación 11. La gran ventaja de esta función criterio radica en la facilidad de cálculo respecto al resto.

4.4.2. A-optimalidad.

Definición 12. El criterio de A-optimización viene definido por la función criterio siguiente:

$$\Phi_A[M(\xi)] = \sum_{i=1}^k \text{var}_\xi \alpha_i = \begin{cases} \text{tr} M^{-1}(\xi) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

La segunda igualdad es consecuencia del hecho de que:

$$\text{var}_\xi \alpha_i = e_i^t M^{-1}(\xi) e_i$$

si los parámetros son estimables.

Proposición 13. Φ_A tiene las siguientes propiedades:

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

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

4.4.3. G-optimalidad.

Definición 14. El criterio de G-optimización viene definido por la función criterio siguiente:

$$\Phi_G[M(\xi)] = \sup_{x \in \mathcal{X}} \text{var}_\xi g_x, \quad \xi \in \Xi$$

Puesto que:

$$g_x(\theta) = \theta(x), \quad x \in \mathcal{X}, \theta \in \Theta$$

entonces:

$$\text{var}_\xi g_x = f^t(x) M^{-1}(\xi) f(x)$$

Cuando $M(\xi)$ es regular. De modo que la función puede escribirse de la forma:

$$\Phi_G[M(\xi)] = \begin{cases} \max_{x \in \mathcal{X}} f^t(x)M^{-1}(\xi)f(x) & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

Proposición 15. Φ_G tiene las siguientes propiedades:

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

4.4.4. E-optimalidad.

Definición 16. El criterio de E-optimización viene definido por la función criterio siguiente:

$$\Phi_E[M(\xi)] = \sup\{\text{var}_\xi \alpha^t c : \|c\| = 1\}, \xi \in \Xi$$

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

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

Proposición 17. Φ_E tiene las siguientes propiedades:

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

4.4.5. c-optimalidad.

Con este criterio el interés se centra en la estimación de combinaciones lineales de los parámetros $c^t\theta$ con mínima varianza. De este modo se da la siguiente definición:

Definición 18. El criterio de c-optimización para un vector c de dimensión k viene definido por la función criterio siguiente:

$$\Phi_D[M(\xi)] = c^t M(\xi) c$$

4.4.6. L-optimalidad.

La siguiente definición es debida a Atwood

Definición 19. El criterio de L-optimización viene definido por la función criterio siguiente:

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

donde W es una matriz definida positiva de orden k .

Proposición 20. Φ_L tiene las siguientes propiedades:

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

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

4.4.7. Φ_p -optimalidad.

La siguiente definición es debida a Atwood

Definición 21. El criterio de Φ_p -optimización es el asociado a la siguiente función criterio:

$$\Phi_p[M(\xi)] = \begin{cases} [k^{-1}\text{tr}M^{-p}(\xi)]^{1/p} & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

siendo p un número real positivo.

4.4.8. L_p -optimalidad.

Este caso viene a generalizar gran parte de los criterios de optimización más usuales:

Definición 22. El criterio de L_p -optimización es asociado a la siguiente función criterio:

$$\Phi_{L_p}[M(\xi)] = \begin{cases} [k^{-1} \text{tr}(HM^{-1}(\xi)H^t)^p]^{1/p} & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases}$$

siendo p un número real positivo y H una matriz regular de orden k .

Proposición 23. Si p es un entero positivo, entonces Φ_A tiene las siguientes propiedades:

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

$$\nabla \Phi_{L_p}[M(\xi)] = \frac{-1}{pm^{1/p}} \left\{ \text{tr}[HM^{-1}(\xi)H^t]^p \right\}^{(1/p)-1} \sum_{k=0}^{p-1} M^{k-p}(\xi) (H^t)^p H^p M^{-k-1}(\xi)$$

4.4.9. I-optimalidad.

Teniendo en cuenta que $f^t(x)M^{-1}(\xi)f(x)$ es la varianza de la predicción de la respuesta en el punto x , este criterio se interesará en minimizar el valor esperado de dichas varianzas sobre el conjunto \mathcal{X} .

Suponiendo que μ es una medida de probabilidad sobre \mathcal{X} se define la función criterio como:

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

4.4.10. MV-optimalidad.

Dicho criterio busca minimizar el máximo de las varianzas de los estimadores de los parámetros. De este modo se centra la atención en las varianzas de los estimadores de todos los parámetros a la vez, sin tener en cuenta las covarianzas. La función criterio será entonces:

$$\Phi_{MV}[M(\xi)] = \begin{cases} \max_i \text{var}_{\xi} \alpha_i & \text{si } \det M(\xi) \neq 0 \\ \infty & \text{si } \det M(\xi) = 0 \end{cases},$$

es decir, se tratará de minimizar el máximo de los elementos de la diagonal de la matriz inversa de

información:

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

4.4.11. Criterio minimax.

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

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

El siguiente teorema muestra la relación entre algunos criterios.

Teorema 24.

1. Si en la L_p -optimización hacemos $H = I$ y hacemos tender a p a infinito obtendremos E -optimización.
2. Si se toma $W = k^{-1}H^tH$ y $p = 1$ la L_p -optimización se transforma en L -optimización.
3. Haciendo $H = I$ y p que tienda a cero la L_p -optimización se convierte en D -optimización.
4. Si en la L -optimización hacemos $W = I$ coincidirá con A -optimización.
5. La Φ_p -optimización es un caso particular de L_p -optimización tomando $H = I$.
6. Llamando $W = \int_{\mathcal{X}} f^t(x)f(x)d\mu$ en L -optimización se tiene I -optimización.
7. Si el criterio minimax se utiliza la norma $|\cdot|_1$ entonces obtenemos el criterio MV , utilizando la norma $|\cdot|_2$ se obtienen el criterio E y utilizando la siguiente norma se obtiene el criterio G :

$$|c|^R = \inf\{\alpha \geq 0 : c \in \alpha R\}$$

donde R es el cierre convexo del conjunto:

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

Demostración. Para demostrar los apartados segundo, cuarto y quinto bastará hacer la sustitución indicada. Para probar el primer apartado denotemos por $\lambda_1, \dots, \lambda_k$ los autovalores de $M(\xi)$ y sea λ_{i_0} el

mínimo de los autovalores. Entonces:

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

Para demostrar el tercer apartado hacemos $H = I$ de modo que

$$\begin{aligned}
\lim_{p \rightarrow 0} \log \left[\frac{1}{k} tr M^{-p}(\xi) \right]^{1/p} &= \lim_{p \rightarrow 0} \frac{\log \frac{1}{k} \sum_{i=1}^k \lambda_i^{-p}}{p} \\
&= -\frac{1}{k} \log \det M(\xi)
\end{aligned}$$

Veamos ahora que se cumple el sexto apartado. En efecto, tomando en L-optimización $W = \int_{\mathcal{X}} f^t(x) f(x) d\mu$ se tendrá

$$\begin{aligned}
tr W M^{-1}(\xi) &= tr \int_{\mathcal{X}} f^t(x) f(x) d\mu M^{-1}(\xi) \\
&= tr \int_{\mathcal{X}} f^t(x) f(x) M^{-1}(\xi) d\mu \\
&= tr \int_{\mathcal{X}} f^t(x) M^{-1}(\xi) f(x) d\mu \\
&= \int_{\mathcal{X}} f^t(x) M^{-1}(\xi) f(x) d\mu
\end{aligned}$$

Para probar el séptimo apartado vemos que

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

siendo e_1, \dots, e_k la base canónica. La segunda igualdad es debida al hecho de que $c^t M^{-1}(\xi) c$ es una función convexa en c y por tanto el máximo se alcanzará en los vértices del conjunto $\{c \in \mathbb{R}^k : |c|_1 = 1\}$,

que vienen dados por los vectores de la base canónica. Por otra parte,

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

Para justificar la primera y segunda igualdades bastará observar que la norma $|\cdot|^R$ caracteriza el conjunto R como la bola unitaria y el borde de la misma será precisamente el conjunto

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

□

4.5. Eficiencia de los diseños.

Observe que los diseños óptimos dependen del criterio de optimalidad usado. Una forma de comparar dos diseños ξ_1, ξ_2 es a partir del cálculo de las ϕ -eficiencias. La ϕ -eficiencia del diseño ξ , abreviada por $\phi\text{-ef}(\xi)$, está definida como:

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

La función $\phi\text{-ef}(\xi)$ está entre cero y uno. Proporciona una medida de cuanto porcentaje en información está suministrando el diseño ξ con respecto a la máxima. Así, si ξ_1 fue obtenido a partir de un criterio de optimalidad ϕ_1 , entonces al calcular la eficiencia de cualquier otro diseño ξ_2 este valor dará una idea de que tan cercano o alejado está éste de la información suministrada por el diseño ϕ_1 -óptimo.

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

$$\begin{aligned}\phi_{c\text{-ef}}(\xi) &= \frac{(c^t M^{-1}(\xi) c)^{-1}}{\sup_{\xi^*} (c^t M^{-1}(\xi^*) c)^{-1}} \\ &= \frac{\inf_{\xi^*} c^t M^{-1}(\xi^*) c}{c^t M^{-1}(\xi) c},\end{aligned}$$

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

En el caso de D-optimalidad:

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

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

Antes de continuar con la teoría concerniente a los diseños óptimos para modelos no lineales, en la siguiente sección se estudian los teoremas de equivalencia.

Capítulo 5

Teorema de Equivalencia.

Kiefer y Wolfowitz demostraron en 1960 la equivalencia de los criterios D-óptimo y G-óptimo. Esto quiere decir que utilizando ambos criterios siempre se llegará al mismo diseño óptimo.

Teorema 25. *Suponiendo que $\sigma(x)$ es una constante para todo x de X , un diseño ξ^* es D-óptimo si, y sólo si es G-óptimo. Es decir los siguientes enunciados son equivalentes:*

1. $\det M(\xi^*) = \max\{\det M(\xi) : \xi \in \Xi\}$.
2. $\max_{x \in X} f^t(x) M^{-1}(\xi^*) f(x) = \min_{\xi \in \Xi} \max_{x \in X} f^t(x) M^{-1}(\xi) f(x)$.

Además, la última expresión es igual a m .

Demostración. 1.- Suponga primero que ξ^* es D-óptimo, es decir que la función Φ_D , alcanza su mínimo en $M(\xi^*)$. Por la proposición (4.10) esta función es finita, continua y diferenciable en el conjunto

$$\{A : A \in \mathbb{R}^{m \times m}, |\log \det A - \log \det M(\xi^*)| < 1\}$$

y

$$\nabla \{-\log \det M(\xi^*)\} = -M^{-1}(\xi^*). \quad (1)$$

Esto sigue que la función

$$\beta \in (0, 1] \mapsto -\log \det [(1 - \beta) M(\xi^*) + \beta f(x) f^t(x)]$$

es finita, continua y diferenciable y alcanza su mínimo en el punto $\beta = 0$. Usando la ecuación (1) obtenemos

$$\begin{aligned}
0 &\leq \left. \frac{\partial \{-\log \det [(1-\beta)M(\xi^*) + \beta f(x)f^t(x)]\}}{\partial \beta} \right|_{\beta=0} \\
&= -\text{tr} [\nabla \log \det M(\xi^*)] [f(x)f^t(x) - M(\xi^*)] \\
&= -f^t(x)M^{-1}(\xi^*)f(x) + m,
\end{aligned}$$

esto es que

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

O de otra forma se tiene que

$$\begin{aligned}
\max_{x \in \mathcal{X}} f^t(x)M^{-1}(\xi)f(x) &\geq \sum_{x \in \mathcal{X}} f^t(x)M^{-1}(\xi)f(x)\xi(x) \\
&= \text{tr}M^{-1}(\xi)M(\xi) \\
&= m,
\end{aligned}$$

para cada diseño no singular ξ . Así se tiene que ξ^* es G-óptimo, y

$$\max_{x \in \mathcal{X}} f^t(x)M^{-1}(\xi^*)f(x) = m. \quad (2)$$

2.- Suponga ahora que ξ^* es G-óptimo.

Se sabe que $M(\xi^*)$ es una matriz no singular. Sea η cualquier diseño D-óptimo. Entonces

$$\begin{aligned}
1 &\leq \det M(\eta)M^{-1}(\xi^*) \\
&= \prod_{i=1}^m \lambda_i, \quad (3)
\end{aligned}$$

donde $\lambda_1, \dots, \lambda_m$ son los eigenvalores de la matriz $M(\eta)M^{-1}(\xi^*)$. Usando la desigualdad bien descono-

cida entre la media aritmética y la media geométrica obtenemos de la ecuación (2)

$$\begin{aligned}
\left(\prod_{i=1}^m \lambda_i \right)^{\frac{1}{m}} &\leq \frac{1}{m} \sum_{i=1}^m \lambda_i \\
&= \frac{1}{m} \text{tr} M(\eta) M^{-1}(\xi^*) \\
&= \frac{1}{m} \sum_{x \in X} f^t(x) M^{-1}(\xi^*) f(x) \eta(x) \\
&\leq \frac{1}{m} \max_{x \in X} f^t(x) M^{-1}(\xi^*) f(x) \\
&= 1. \tag{4}
\end{aligned}$$

Comparando las ecuaciones (3) y (4) se obtiene

$$\det M(\xi^*) = \det M(\eta),$$

así ξ^* es un diseño D-óptimo. □

Observación 26. 1.- Si $\sigma(x)$ no es constante, el teorema de equivalencia se cumple también haciendo una pequeña precisión. En este caso puede demostrarse que ξ^* es D-óptimo si, y sólo si

$$\max_{x \in X} \sigma^{-2}(x) f^t(x) M^{-1}(\xi^*) f(x) = \min_{\xi \in \Xi} \max_{x \in X} \sigma^{-2} f^t(x) M^{-1}(\xi) f(x).$$

siendo además la última expresión igual a m .

5.1. Teorema General de Equivalencia.

En 1973 Whittle y en 1974 Kiefer J. generalizan el teorema de equivalencia para funciones criterio más generales. Con el objeto de enunciar estos resultados damos una definición a o más rigurosa de función criterio. Para ello nos fijaremos en las propiedades que verificaban las funciones criterio vistas anteriormente. Observamos que todas las funciones criterio vistas son convexas y continuas. Con el objeto de salvaguardar estas propiedades daremos la siguiente definición:

Definición 27. Diremos que una función real Φ en las condiciones de las funciones criterio, según la definición dada más arriba, es una función criterio convexa si cumple las propiedades:

1. Existe U_Φ abierto de $\mathcal{L}(\mathcal{M})$ tal que $U_\Phi \supset \mathcal{M}_+$ y Φ está definida, es finita y convexa en U_Φ .

2. Si:

$$M_n \in \mathcal{M}_+, n = 1, 2, \dots \text{ y } \lim_{n \rightarrow \infty} M_n = M \in \mathcal{M} - \mathcal{M}_+$$

entonces:

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

Observación 28. No es necesario que U_Φ sea convexo. Entenderemos entonces que Φ debe ser convexa en los subconjuntos convexos de U_Φ .

Proposición 29. *Si U es un abierto de $\mathcal{L}(\mathcal{M})$ y Φ es convexa y finita en U , entonces Φ es continua en U .*

Demostración. Denote la dimensión de $\mathcal{L}(\mathcal{M})$ por r . Escogiendo $L_1 \in U$. Puesto que U es abierto hay una esfera abierta r -dimensional

$$\mathcal{G} = \{L : L \in \mathcal{L}(\mathcal{M}), \|L - L_1\| < \delta\}$$

y un simplex r -dimensional \mathcal{J} así que $\mathcal{G} \subset \mathcal{J} \subset U$. Denote los vertices del simplex \mathcal{J} por V_1, \dots, V_r . De la convexidad de Φ resulta que

$$\Phi(L) \leq \max \{\Phi(V_i) : i = 0, \dots, r\} = d < \infty; \quad (L \in \mathcal{G}).$$

Tomando $\varepsilon \in (0, 1)$ y denotando

$$\mathcal{G}_\varepsilon = \{L : L \in \mathcal{L}(\mathcal{M}), \|L - L_1\| < \varepsilon\delta\}.$$

Para $L \in \mathcal{G}_\varepsilon$ defina $W = (L - L_1)/\varepsilon$, tal que

$$L = (1 - \varepsilon)L_1 + \varepsilon(L_1 + W)$$

y la convexidad de Φ implica

$$\Phi(L) \leq (1 - \varepsilon)\Phi(L_1) + \varepsilon d. \quad (5)$$

Por otra parte, se tiene la igualdad

$$L = \frac{1}{1+\varepsilon}L + \left(1 - \frac{1}{1+\varepsilon}\right)(L_1 - W).$$

Por lo tanto, usando la convexidad de Φ otra vez, se tiene

$$\Phi(L_1) \leq \frac{1}{1+\varepsilon}\Phi(L) + \frac{\varepsilon d}{1+\varepsilon}. \quad (6)$$

De las ecuaciones (5) y (6) resulta que para cada $\varepsilon \in (0, 1)$

$$|\Phi(L) - \Phi(L_1)| \leq \varepsilon[d - \Phi(L_1)]$$

cuando $L \in \mathcal{G}_\varepsilon$. La continuidad requerida de Φ en L_1 se sigue inmediatamente. □

Proposición 30. *Cualquier función criterio convexa, con la definición dada, es continua en \mathcal{M} .*

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

Todavía necesitamos dar una definición de derivada direccional debida a Frechét.

Definición 31. Dada una función real Φ , convexa y definida en un subconjunto convexo de un espacio euclídeo, y dados dos puntos en ese conjunto x y v , se define la derivada direccional de Φ en el punto x y en la dirección de v como

$$\partial\Phi(x, v) = \lim_{\beta \rightarrow 0^+} \frac{\Phi[(1-\beta)x + \beta v] - \Phi(x)}{\beta}.$$

Observación 32. La derivada direccional existe siempre gracias a la convexidad de Φ . En efecto, demostraremos que la función:

$$\varphi : \beta \in (0, 1) \rightarrow \frac{\Phi[(1-\beta)x + \beta v] - \Phi(x)}{\beta}$$

es creciente en $(0, 1)$. Por tanto el límite siguiente siempre existe, siendo finito o menos infinito

$$\lim_{\beta \rightarrow 0^+} \frac{\Phi[(1-\beta)x + \beta v] - \Phi(x)}{\beta}.$$

Veamos que φ es una función creciente. Para ello tomando $0 < \beta_1 < \beta_2 < 1$ y calculamos:

$$\begin{aligned}\Phi[(1 - \beta_1)x + \beta_1 v] - \Phi(x) &= \Phi\left\{\frac{\beta_1}{\beta_2}[(1 - \beta_2)x + \beta_2 v] + \left(1 - \frac{\beta_1}{\beta_2}\right)x\right\} - \Phi(x) \\ &\leq \frac{\beta_1}{\beta_2}\Phi[(1 - \beta_2)x + \beta_2 v] + \left(1 - \frac{\beta_1}{\beta_2}\right)\Phi(x) - \Phi(x) \\ &= \frac{\beta_1}{\beta_2}\{\Phi[(1 - \beta_2)x + \beta_2 v] - \Phi(x)\}\end{aligned}$$

Pueden darse otras definiciones de derivada direccional, pero nos interesará la anterior dado al uso de las funciones convexas. Si existe el gradiente de Φ entonces por su definición, podemos escribir:

$$\begin{aligned}\partial\Phi(x, v) &= \text{tr}\{[\nabla\Phi(x)](v - x)\} \\ &= \langle \nabla\Phi(x), v - x \rangle.\end{aligned}$$

Recíprocamente, se puede definir el gradiente como aquella matriz que encaja en la expresión anterior.

Proposición 33. *Si Φ es finito y convexo en un subconjunto abierto de U de \mathcal{L} y si $\nabla_L\Phi(L|\mathcal{L})$ es definido en cualquier $L \in \mathcal{L}$, entonces el mapeo*

$$L \in U \mapsto \nabla_L\Phi(L|\mathcal{L})$$

es continuo.

Teorema 34. *(Teorema general de equivalencia) Sea Φ una función criterio convexa y ξ^* un diseño tal que*

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

Entonces son equivalentes:

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

$$\varphi : \beta \in [0, 1] \mapsto \Phi[(1 - \beta)M(\xi^*) + \beta M(\xi)]$$

tiene un mínimo local en $\beta = 0$.

3. $\partial\Phi[M(\xi^*), M(\xi)] \geq 0$, $\xi \in \Xi$

Demostración. Es obvio que 1 implica 2. Si ξ^* es Φ -óptimo local, entonces para cada $\xi \in \Xi$ hay $\delta > 0$, tal que

$$\frac{\Phi [(1 - \beta) M (\xi^*) + \beta M (\xi)] - \Phi [M (\xi^*)]}{\beta} \geq 0$$

para cada $\beta \in (0, \delta)$. Resulta que $\partial\Phi [M (\xi^*), M (\xi)] \geq 0$, es decir, $2 \Rightarrow 3$ se ha demostrado.

Para probar $3 \Rightarrow 1$ supongase que ξ^* no Φ -óptimo. Entonces hay un diseño ξ talque $\Phi [M (\xi)] < \Phi [M (\xi^*)]$. Por lo tanto para cada $\beta \in (0, 1)$

$$\begin{aligned} \frac{\Phi [(1 - \beta) M (\xi^*) + \beta M (\xi)] - \Phi [M (\xi^*)]}{\beta} &\leq \frac{(1 - \beta) \Phi [M (\xi^*)] + \beta \Phi [M (\xi)] - \Phi [M (\xi^*)]}{\beta} \\ &= \Phi [M (\xi)] - \Phi [M (\xi^*)], \end{aligned}$$

así $\partial\Phi [M (\xi^*), M (\xi)] \leq \Phi [M (\xi)] - \Phi [M (\xi^*)] < 0$, por lo que la desigualdad en 3) no se cumple. \square

Obsérvese que este teorema aporta un criterio general (tercer apartado) para contrastar si un diseño dado es o no Φ -óptimo, tanto si la función criterio es diferenciable como si no lo es. Cuando n o la función Φ sea diferenciable, se puede dar el siguiente teorema:

Teorema 35. *Si Φ es diferenciable en un entorno de $M (\xi^*)$, entonces son equivalentes:*

1. ξ^* es Φ -óptimo.

$$2. f^t(x) \nabla \Phi [M (\xi^*)] f(x) \geq \text{tr} M (\xi^*) \nabla \Phi [M (\xi^*)], \quad x \in X. \quad (7)$$

$$3. \min_{x' \in X} f^t(x') \nabla \Phi [M (\xi^*)] f(x') = \sum_{x \in X} f^t(x) \nabla \Phi [M (\xi^*)] f(x) \xi^*(x). \quad (8)$$

Demostración. De la definición de derivada direccional de Φ se tiene

$$\partial\Phi [M (\xi^*), f(x) f^t(x)] = f^t(x) \nabla \Phi [M (\xi^*)] f(x) - \text{tr} M (\xi^*) \nabla \Phi [M (\xi^*)].$$

Por lo tanto la implicación $1 \Rightarrow 2$ es una consecuencia de la última proposición así $f(x) f^t(x) = M (\xi_x)$.

Si, a su vez, la ecuación (7) es multiplicada por $\xi(x)$, y el resultado es sumado bajo X , entonces

$$\partial\Phi [M (\xi^*), M (\xi)] \geq 0; \quad (\xi \in \Xi).$$

Por lo tanto, según la última proposición, ξ^* es Φ -óptimo. La implicación $3 \Rightarrow 2$ se sigue claramente

de

$$\text{tr} M(\xi^*) \nabla \Phi [M(\xi^*)] = \sum_{x \in X} \text{tr} f(x) f^t(x) \nabla \Phi [M(\xi^*) \xi^*(x).]$$

A la inversa de 2) y de la desigualdad

$$\min_{x' \in X} f^t(x') \nabla \Phi [M(\xi^*)] f(x') \leq \sum_{x \in X} f^t(x) \nabla \Phi [M(\xi^*)] f(x) \xi^*(x),$$

se obtiene el enunciado en 3). □

Teorema. *En un diseño D-óptimo los pesos son uniformes.*

Demostración. Supongamos que la medida ξ concentra su masa en w_1, w_2, \dots, w_n ($\sum_{i=1}^n w_i = 1$), en n puntos distintos x_1, x_2, \dots, x_n , respectivamente. Para esta elección de ξ los elementos de $M(\xi)$ son

$$m_{i,j} = \sum_{l=1}^n w_l f_i(x_l) f_j(x_l)$$

haciendo $a_{il} = w_l f_i(x_l)$ y $b_{jl} = f_j(x_l)$ para $i, j, l = 1, 2, \dots, n$. Se deduce que el determinante toma la forma

$$\left\{ \prod_{l=1}^n w_l \right\} \left\{ \det \|f_i(x_j)\|_{i,j=0}^n \right\}^2$$

si las f_1, f_2, \dots, f_n son especificadas como

$$f_i(x) = x^i \lambda^{1/2}(x), \quad i = 1, 2, \dots, n, \quad x \in [a, b],$$

entonces se tiene

$$\prod_{l=1}^n w_l \prod_{l=1}^n \lambda(x_l) \prod_{0 \leq i < j \leq n} (x_i - x_j)^2$$

maximizando con respecto a w 's se tiene

$$\ell = \sum_{l=1}^n \log w_l + \sum_{l=1}^n \log \lambda(x_l) + \sum_{0 \leq i < j \leq n} 2 \log (x_i - x_j)$$

para ello utilizaremos multiplicadores de Lagrange

$$\ell = \sum_{l=1}^n \log w_l + \sum_{l=1}^n \log \lambda(x_l) + \sum_{0 \leq i < j \leq n} 2 \log (x_i - x_j) - \alpha \left(\sum_{l=1}^n w_l - 1 \right)$$

derivando con respecto a w_l obtenemos

$$\frac{\partial \ell}{\partial w_l} = \frac{1}{w_l} - \alpha,$$

igualando a cero y despejando w_l se llega a que

$$w_l = \frac{1}{\alpha},$$

recordando que $\sum_{l=1}^n w_l = 1$ se tiene que $\alpha = n$. Por lo tanto $w_l = \frac{1}{n}$. □

Yang y Stufken (2009) proponen un nuevo enfoque para identificar los puntos de soporte para un diseño localmente óptimo cuando el modelo es un modelo no lineal. En contraste con el enfoque geométrico comúnmente utilizado, usaron un enfoque basado en herramientas algebraicas. Consideraron limitaciones a modelos con dos parámetros, y llegaron a la conclusión para el modelo de Michaelis-Menten solo es necesario dos puntos de soporte.

Capítulo 6

Modelo de Michaelis-Menten.

6.1. Introducción.

Muchos modelos estadísticos comunes pueden ser expresados como un modelo no lineal, que son parámetros asociados con una población entera o con ciertos niveles repetibles de factores experimentales. Un ejemplo de modelos no lineales son los modelos racionales en donde le daremos más énfasis al modelo de Michaelis-Menten para la cinética de la enzima, que relaciona la velocidad inicial de reacción de la enzima para la concentración de sustrato el cual está constituido por dos parámetros. Es el modelo no lineal que se describe brevemente en el capítulo 2. Por ejemplo, el diseño para el modelo propuesto por Michaelis y Menten (1913) se discute por Dette et al. (2003), Dette y Biedermann (2003) y Dette et al. (2005). El diseño óptimo de experimentos cuando los modelos son exponenciales se investigó por Mukhopadhyay y Haines (1995), Dette y Neugebauer (1997), y Han y Chaloner (2003).

6.2. Modelo de Michaelis-Menten.

Supongamos que tenemos n observaciones (y_i, x_i) , $i = 1, \dots, n$, se encuentran disponibles en la tasa observada de la reacción y la concentración de sustrato correspondiente, entonces el modelo de Michaelis-Menten puede ser escrito como

$$\begin{aligned} y_i &= \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i \\ &= \eta(x_i, \theta) + \varepsilon_i, \end{aligned} \tag{6.1}$$

Donde $\eta(x_i, \theta) = \theta_1 x_i / \theta_2 + x_i$, $\theta = (\theta_1, \theta_2)'$ y $\varepsilon_i \sim N(0, \sigma^2)$. Por lo tanto $y_i \sim N(\eta(x_i, \theta), \sigma^2)$.

Ahora, la función de verosimilitud para un punto del soporte y_i , $i = 1, \dots, n$ puede ser escrita como,

$$\begin{aligned} L(\theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} [y_i - \eta(x_i, \theta)]^2\right\} \\ &= (2\pi\sigma^2)^{1/2} \exp\left\{-\frac{1}{2\sigma^2} [y_i - \eta(x_i, \theta)]^2\right\}. \end{aligned} \quad (6.2)$$

La log verosimilitud puede ser escrita como,

$$l(\theta) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} [y_i - \eta(x_i, \theta)]^2. \quad (6.3)$$

Utilizando la ecuación (6.3) se construye la matriz de información de Fisher para encontrar el diseño óptimo.

6.3. Matriz de Información de Fisher.

Para un modelo que implica el vector de parámetros θ y un diseño de medida ξ_n en la ecuación (4.2) en el intervalo $(0, x_0)$ con puntos de diseño finitos, donde los puntos de diseño finitos corresponden a la ubicación de las observaciones y los pesos dan las proporciones relativas del total de observaciones que deben adoptarse en los puntos correspondientes. A continuación, la matriz de información de Fisher se basa en los valores esperados de las segundas derivadas parciales, y está dada por

$$M(\theta, \xi_n) = E\left[-\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}\right], \quad (6.4)$$

donde $j = 1, 2$ y $l(\theta)$ es la función de log verosimilitud definida en la ecuación (6.3).

Las primeras derivadas parciales de la log verosimilitud son

$$\begin{aligned} \frac{\partial l(\theta)}{\partial\theta_1} &= \frac{1}{\sigma^2} [y_i - \eta(x_i, \theta)] \frac{x_i}{\theta_2 + x_i}, \\ \frac{\partial l(\theta)}{\partial\theta_2} &= -\frac{1}{\sigma^2} [y_i - \eta(x_i, \theta)] \frac{\theta_1 x_i}{(\theta_2 + x_i)^2}. \end{aligned}$$

Tomando esperanza en la segunda derivada parcial de la log verosimilitud con respecto a todos los parámetros y multiplicando por -1 daría la matriz de información de Fisher 2×2 para los parámetros

θ_1 y θ_2 .

$$\begin{aligned}
-E \left[\frac{\partial^2 l(\theta)}{\partial \theta_1^2} \right] &= -E \left[\frac{1}{\sigma^2} [y_i - \eta(x_i, \theta)] \frac{\partial}{\partial \theta_1} \left(\frac{x_i}{\theta_2 + x_i} \right) - \frac{\partial}{\partial \theta_1} \eta(x_i, \theta) \frac{x_i}{\theta_2 + x_i} \right] \\
&= \frac{1}{\sigma^2} \frac{x_i}{\theta_2 + x_i} \frac{x_i}{\theta_2 + x_i} \\
&= \frac{1}{\sigma^2} \frac{x_i^2}{(\theta_2 + x_i)^2}.
\end{aligned}$$

$$\begin{aligned}
-E \left[\frac{\partial^2 l(\theta)}{\partial \theta_2^2} \right] &= -E \left[\frac{1}{\sigma^2} [y_i - \eta(x_i, \theta)] \frac{\partial}{\partial \theta_2} \left(\frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right) - \frac{\partial}{\partial \theta_2} \eta(x_i, \theta) \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \right] \\
&= \frac{1}{\sigma^2} \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \\
&= \frac{1}{\sigma^2} \frac{\theta_1^2 x_i^2}{(\theta_2 + x_i)^4}.
\end{aligned}$$

$$\begin{aligned}
-E \left[\frac{\partial^2 l(\theta)}{\partial \theta_1 \partial \theta_2} \right] &= -E \left[\frac{1}{\sigma^2} [y_i - \eta(x_i, \theta)] \frac{\partial}{\partial \theta_2} \left(\frac{x_i}{\theta_2 + x_i} \right) - \frac{\partial}{\partial \theta_2} \eta(x_i, \theta) \frac{x_i}{\theta_2 + x_i} \right] \\
&= -\frac{1}{\sigma^2} \frac{\theta_1 x_i}{(\theta_2 + x_i)^2} \frac{x_i}{\theta_2 + x_i} \\
&= -\frac{1}{\sigma^2} \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3}.
\end{aligned}$$

Entonces la matriz de información puede ser escrita como

$$\begin{aligned}
M(\xi) &= \sum_{x_1, x_2} \begin{pmatrix} \frac{x_i^2}{(\theta_2 + x_i)^2} & -\frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} \\ -\frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} & \frac{\theta_1^2 x_i^2}{(\theta_2 + x_i)^4} \end{pmatrix} w_i \\
&= \begin{pmatrix} \sum_{x_1, x_2} \frac{x_i^2}{(\theta_2 + x_i)^2} w_i & -\sum_{x_1, x_2} \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} w_i \\ -\sum_{x_1, x_2} \frac{\theta_1 x_i^2}{(\theta_2 + x_i)^3} w_i & \sum_{x_1, x_2} \frac{\theta_1^2 x_i^2}{(\theta_2 + x_i)^4} w_i \end{pmatrix} \\
&= \begin{pmatrix} \frac{x_1^2 w_1}{(\theta_2 + x_1)^2} + \frac{x_2^2 w_2}{(\theta_2 + x_2)^2} & -\frac{\theta_1 x_1^2 w_1}{(\theta_2 + x_1)^3} - \frac{\theta_1 x_2^2 w_2}{(\theta_2 + x_2)^3} \\ -\frac{\theta_1 x_1^2 w_1}{(\theta_2 + x_1)^3} - \frac{\theta_1 x_2 w_2}{(\theta_2 + x_2)^3} & \frac{\theta_1^2 x_1^2 w_1}{(\theta_2 + x_1)^4} + \frac{\theta_1^2 x_2^2 w_2}{(\theta_2 + x_2)^4} \end{pmatrix} \\
&= \begin{pmatrix} \frac{x_1^2 w_1 (\theta_2 + x_2)^2 + x_2^2 w_2 (\theta_2 + x_1)^2}{(\theta_2 + x_1)^2 (\theta_2 + x_2)^2} & -\theta_1 \left[\frac{x_1^2 w_1 (\theta_2 + x_2)^3 + x_2^2 w_2 (\theta_2 + x_1)^3}{(\theta_2 + x_1)^3 (\theta_2 + x_2)^3} \right] \\ -\theta_1 \left[\frac{x_1^2 w_1 (\theta_2 + x_2)^3 + x_2^2 w_2 (\theta_2 + x_1)^3}{(\theta_2 + x_1)^3 (\theta_2 + x_2)^3} \right] & \theta_1^2 \left[\frac{x_1^2 w_1 (\theta_2 + x_2)^4 + x_2^2 w_2 (\theta_2 + x_1)^4}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \right] \end{pmatrix}, \quad (6.5)
\end{aligned}$$

ya que $M(\xi)$ es una combinación ponderada de las matrices de información de Fisher asociada a cada

punto de soporte x_i del diseño. La matriz de la información dependen de los parámetros desconocidos, por lo tanto, el diseño óptimo depende sobre los parámetros desconocidos y este diseño óptimo es conocido como el diseño óptimo local (Atkinson et al., 2007).

6.4. Diseños Localmente D-óptimos.

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

Suponiendo que el modelo (6.1) y el diseño ξ_n en la ecuación (4.2), la matriz de información de Fisher esperada de θ en la ecuación (5.5) y el diseño D-óptimo local maximiza el determinante de la matriz de información a partir de (4.9). El determinante de la matriz de información es:

$$\begin{aligned}
|M(\xi)| &= \left\{ \frac{x_1^2 w_1 (\theta_2 + x_2)^2 + x_2^2 w_2 (\theta_2 + x_1)^2}{(\theta_2 + x_1)^2 (\theta_2 + x_2)^2} \right\} \left\{ \theta_1^2 \left[\frac{x_1^2 w_1 (\theta_2 + x_2)^4 + x_2^2 w_2 (\theta_2 + x_1)^4}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \right] \right\} \\
&- \left\{ -\theta_1 \left[\frac{x_1^2 w_1 (\theta_2 + x_2)^3 + x_2^2 w_2 (\theta_2 + x_1)^3}{(\theta_2 + x_1)^3 (\theta_2 + x_2)^3} \right] \right\}^2 \\
&= \theta_1^2 \left[\frac{x_1^4 w_1^2 (\theta_2 + x_2)^6 + x_1^2 x_2^2 w_1 w_2 (\theta_2 + x_1)^2 (\theta_2 + x_2)^2 + x_1^2 x_2^2 w_1 w_2 (\theta_2 + x_1)^2 (\theta_2 + x_2)^4 + x_2^4 w_2^2 (\theta_2 + x_1)^6}{(\theta_2 + x_1)^6 (\theta_2 + x_2)^6} \right] \\
&- \theta_1^2 \left[\frac{x_1^4 w_1^2 (\theta_2 + x_2)^6 + x_1^2 x_2^2 w_1 w_2 (\theta_2 + x_1)^3 (\theta_2 + x_2)^3 + x_1^2 x_2^2 w_1 w_2 (\theta_2 + x_1)^3 (\theta_2 + x_2)^3 + x_2^4 w_2^2 (\theta_2 + x_1)^6}{(\theta_2 + x_1)^6 (\theta_2 + x_2)^6} \right] \\
&= \theta_1^2 \left[\frac{x_1^2 x_2^2 w_1 w_2 (\theta_2 + x_1)^4 (\theta_2 + x_2)^2 + x_1^2 x_2^2 w_1 w_2 (\theta_2 + x_1)^2 (\theta_2 + x_2)^4 - 2x_1^2 x_2^2 w_1 w_2 (\theta_2 + x_1)^3 (\theta_2 + x_2)^3}{(\theta_2 + x_1)^6 (\theta_2 + x_2)^6} \right] \\
&= \theta_1^2 \left\{ x_1^2 x_2^2 w_1 w_2 (\theta_2 + x_1)^4 (\theta_2 + x_2)^2 \left[\frac{(\theta_2 + x_1)^2 - 2(\theta_2 + x_1)(\theta_2 + x_2) + (\theta_2 + x_2)^2}{(\theta_2 + x_1)^6 (\theta_2 + x_2)^6} \right] \right\} \\
&= \theta_1^2 \left\{ \frac{x_1^2 x_2^2 w_1 w_2 [(\theta_2 + x_1) - (\theta_2 + x_2)]^2}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \right\} \\
&= \frac{\theta_1^2 x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \tag{6.6}
\end{aligned}$$

Si $x_1 < x_2$ esta función es creciente en x_2 en $[L, R]$, así $x_2 = R$. Con esta opción el determinante se convierte en

$$T(x) = \frac{\theta_1^2 x^2 w_1 (x - R)^2 R(1 - w_1)}{(\theta_2 + x)^4 (\theta_2 + R)^4}, \tag{6.7}$$

donde $x = x_1$, $w_2 = 1 - w_1$. Ahora se realiza la maximización de la función criterio,

$$\begin{aligned}
-\log |M(\xi)| &= -\log T(x) \\
&= -\log \left(\frac{\theta_1^2 x^2 w_1 (x - R)^2 R (1 - w_1)}{(\theta_2 + x)^4 (\theta_2 + R)^4} \right) \\
&= -\log \theta_1^2 - 2 \log x - \log w_1 - 2 \log(x - R) - \log R^2 \\
&\quad - \log(1 - w_1) + 4 \log(\theta_2 + x) + \log(\theta_2 + R)^4
\end{aligned} \tag{6.8}$$

Ahora para determinar los valores de x y w_1 se derivará con respecto a x y a w_1 , como se muestra a continuación:

$$\begin{aligned}
\frac{-\partial \log |M(\xi)|}{\partial x} &= -\frac{2}{x} - \frac{2}{x - R} + \frac{4}{\theta_2 + x} \\
&= \frac{-2(x - R)(\theta_2 + x) - 2x(\theta_2 + x) + 4x(x - R)}{x(x - R)(\theta_2 + x)} \\
&= \frac{-2(x^2 + x\theta_2 - Rx - R\theta_2) - 2(x\theta_2 + x^2) + 4(x^2 - Rx)}{x(x - R)(\theta_2 + x)} \\
&= \frac{-2x^2 - 2x\theta_2 + 2Rx + 2R\theta_2 - 2x\theta_2 - 2x^2 + 4x^2 - 4Rx}{x(x - R)(\theta_2 + x)} \\
&= \frac{-4x\theta_2 + 2R\theta_2 - 2Rx}{x(x - R)(\theta_2 + x)},
\end{aligned} \tag{6.9}$$

Por otro lado tenemos

$$\begin{aligned}
\frac{-\partial \log |M(\xi)|}{\partial w_1} &= -\frac{1}{w_1} + \frac{1}{1 - w_1} \\
&= \frac{-1 + w_1 + w_1}{w_1(1 - w_1)} \\
&= \frac{1 - 2w_1}{w_1(1 - w_1)}
\end{aligned} \tag{6.10}$$

Igualando a cero y despejando x y w_1 de (6.9) y (6.10) respectivamente, se obtiene

$$x = \frac{R\theta_2}{2\theta_2 + R}, \quad w_1 = \frac{1}{2},$$

por lo tanto el diseño D-óptimo para el modelo de Michaelis-Menten es

$$\xi^* = \begin{pmatrix} \frac{R\theta_2}{2\theta_2 + R} & R \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Veamos que se trata realmente del diseno D-ptimo utilizando para ello el teorema de equivalencia:

$$\begin{aligned}
M^{-1}(\xi) &= \begin{pmatrix} \theta_1^2 \left[\frac{x_1^2 w_1 (\theta_2 + x_2)^4 + x_2^2 w_2 (\theta_2 + x_1)^4}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \right] & \theta_1 \left[\frac{x_1^2 w_1 (\theta_2 + x_2)^3 + x_2^2 w_2 (\theta_2 + x_1)^3}{(\theta_2 + x_1)^3 (\theta_2 + x_2)^3} \right] \\ \theta_1^2 \left[\frac{x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \right] & \theta_1^2 \left[\frac{x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \right] \\ \theta_1 \left[\frac{x_1^2 w_1 (\theta_2 + x_2)^3 + x_2^2 w_2 (\theta_2 + x_1)^3}{(\theta_2 + x_1)^3 (\theta_2 + x_2)^3} \right] & \frac{x_1^2 w_1 (\theta_2 + x_2)^2 + x_2^2 w_2 (\theta_2 + x_1)^2}{(\theta_2 + x_1)^2 (\theta_2 + x_2)^2} \\ \theta_1^2 \left[\frac{x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \right] & \theta_1^2 \left[\frac{x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2}{(\theta_2 + x_1)^4 (\theta_2 + x_2)^4} \right] \end{pmatrix} \\
&= \begin{pmatrix} \frac{x_1^2 w_1 (\theta_2 + x_2)^4 + x_2^2 w_2 (\theta_2 + x_1)^4}{x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2} & \frac{[x_1^2 w_1 (\theta_2 + x_2)^3 + x_2^2 w_2 (\theta_2 + x_1)^3](\theta_2 + x_1)(\theta_2 + x_2)}{\theta_1 x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2} \\ \frac{[x_1^2 w_1 (\theta_2 + x_2)^3 + x_2^2 w_2 (\theta_2 + x_1)^3](\theta_2 + x_1)(\theta_2 + x_2)}{\theta_1 x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2} & \frac{[x_1^2 w_1 (\theta_2 + x_2)^2 + x_2^2 w_2 (\theta_2 + x_1)^2](\theta_2 + x_1)^2 (\theta_2 + x_2)^2}{\theta_1^2 x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2} \end{pmatrix} \\
&= \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}
\end{aligned}$$

veamos a que se reduce cada entrada de esta matriz sustituyendo los valores de las x_i 's y w_i 's

$$\begin{aligned}
x_1^2 w_1 (x_1 - x_2)^2 x_2^2 w_2 &= \left(\frac{R\theta_2}{2\theta_2 + R} \right)^2 \left(\frac{1}{2} \right) \left(\frac{R\theta_2}{2\theta_2 + R} - R \right)^2 R^2 \left(\frac{1}{2} \right) \\
&= \left(\frac{1}{4} \right) \left(\frac{R^2 \theta_2}{2\theta_2 + R} \right)^2 \left(\frac{-R\theta_2 - R^2}{2\theta_2 + R} \right)^2 \\
&= \left(\frac{1}{4} \right) \left(\frac{-R^2 \theta_2 (R\theta_2 + R^2)}{(2\theta_2 + R)^2} \right)^2 \\
&= \left(\frac{1}{4} \right) \left(\frac{R^6 \theta_2^2 (\theta_2 + R)^2}{(2\theta_2 + R)^4} \right)
\end{aligned}$$

$$\begin{aligned}
x_1^2 w_1 (\theta_2 + x_2)^4 + x_2^2 w_2 (\theta_2 + x_1)^4 &= \left(\frac{R\theta_2}{2\theta_2 + R} \right)^2 \left(\frac{1}{2} \right) (\theta_2 + R)^4 + R^2 \left(\frac{1}{2} \right) \left(\theta_2 + \frac{R\theta_2}{2\theta_2 + R} \right)^4 \\
&= \frac{R^2 \theta_2^2 (\theta_2 + R)^4}{2 (2\theta_2 + R)^2} + \frac{R^2 (2\theta_2^2 + 2R\theta_2)^4}{2 (2\theta_2 + R)^4} \\
&= \frac{R^2 \theta_2^2 (\theta_2 + R)^4}{2 (2\theta_2 + R)^2} + \frac{R^2 (2\theta_2)^4 (\theta_2 + R)^4}{2 (2\theta_2 + R)^4} \\
&= \frac{R^2 \theta_2^2 (\theta_2 + R)^4 (2\theta_2 + R)^2 + R^2 (2\theta_2)^4 (\theta_2 + R)^4}{2 (2\theta_2 + R)^4} \\
&= \frac{R^2 \theta_2^2 (\theta_2 + R)^4}{2 (2\theta_2 + R)^4} \left[(2\theta_2 + R)^2 + 16\theta_2^2 \right]
\end{aligned}$$

$$\begin{aligned}
\left[x_1^2 w_1 (\theta_2 + x_2)^3 + x_2^2 w_2 (\theta_2 + x_1)^3 \right] (\theta_2 + x_1) (\theta_2 + x_2) &= \left[\left(\frac{R\theta_2}{2\theta_2 + R} \right)^2 \left(\frac{1}{2} \right) (\theta_2 + R)^3 + R^2 \left(\frac{1}{2} \right) \left(\theta_2 + \frac{R\theta_2}{2\theta_2 + R} \right)^3 \right] \\
&\times \left(\theta_2 + \frac{R\theta_2}{2\theta_2 + R} \right) (\theta_2 + R) \\
&= \left[\frac{R^2 \theta_2^2 (\theta_2 + R)^3}{2 (2\theta_2 + R)^2} + \frac{R^2 (2\theta_2^2 + 2R\theta_2)^3}{2 (2\theta_2 + R)^3} \right] \\
&\times \left(\frac{2\theta_2^2 + 2R\theta_2}{2\theta_2 + R} \right) (\theta_2 + R) \\
&= \left[\frac{R^2 \theta_2^2 (\theta_2 + R)^3}{2 (2\theta_2 + R)^2} + \frac{R^2 (2\theta_2)^3 (\theta_2 + R)^3}{2 (2\theta_2 + R)^3} \right] \left(\frac{2\theta_2 (\theta_2 + R)}{2\theta_2 + R} \right) (\theta_2 + R) \\
&= \left[\frac{R^2 \theta_2^2 (\theta_2 + R)^3 (2\theta_2 + R)^3 + R^2 (2\theta_2)^3 (\theta_2 + R)^3}{2 (2\theta_2 + R)^3} \right] \left(\frac{2\theta_2 (\theta_2 + R)^2}{2\theta_2 + R} \right) \\
&= \left[\frac{R^2 \theta_2^2 (\theta_2 + R)^3}{2 (2\theta_2 + R)^3} [(2\theta_2 + R) + 8\theta_2] \right] \left(\frac{2\theta_2 (\theta_2 + R)^2}{2\theta_2 + R} \right) \\
&= \frac{R^2 \theta_2^3 (\theta_2 + R)^5 (10\theta_2 + R)}{(2\theta_2 + R)^4}
\end{aligned}$$

$$\begin{aligned}
\left[x_1^2 w_1 (\theta_2 + x_2)^2 + x_2^2 w_2 (\theta_2 + x_1)^2 \right] (\theta_2 + x_1)^2 (\theta_2 + x_2)^2 &= \left[\left(\frac{R\theta_2}{2\theta_2 + R} \right)^2 \left(\frac{1}{2} \right) (\theta_2 + R)^2 + R^2 \left(\frac{1}{2} \right) \left(\theta_2 + \frac{R\theta_2}{2\theta_2 + R} \right)^2 \right] \\
&\times \left(\theta_2 + \frac{R\theta_2}{2\theta_2 + R} \right)^2 (\theta_2 + R)^2 \\
&= \left[\frac{R^2 \theta_2^2 (\theta_2 + R)^2}{2 (2\theta_2 + R)^2} + \frac{R^2 (2\theta_2^2 + 2R\theta_2)^2}{2 (2\theta_2 + R)^2} \right] \\
&\times \left(\frac{2\theta_2^2 + 2R\theta_2}{2\theta_2 + R} \right)^2 (\theta_2 + R)^2 \\
&= \left[\frac{R^2 \theta_2^2 (\theta_2 + R)^2}{2 (2\theta_2 + R)^2} + \frac{R^2 (2\theta_2)^2 (\theta_2 + R)^2}{2 (2\theta_2 + R)^2} \right] \left(\frac{2\theta_2 (\theta_2 + R)}{2\theta_2 + R} \right)^2 (\theta_2 + R)^2 \\
&= \left[\frac{R^2 \theta_2^2 (\theta_2 + R)^2 + R^2 (2\theta_2)^2 (\theta_2 + R)^2}{2 (2\theta_2 + R)^2} \right] \left(\frac{(2\theta_2)^2 (\theta_2 + R)^4}{(2\theta_2 + R)^2} \right) \\
&= \left[\frac{20R^2 \theta_2^2 (\theta_2 + R)^3}{2 (2\theta_2 + R)^3} \right] \left(\frac{4\theta_2^2 (\theta_2 + R)^4}{2\theta_2 + R} \right) \\
&= \frac{10R^2 \theta_2^4 (\theta_2 + R)^6}{(2\theta_2 + R)^4}
\end{aligned}$$

asi llegamos a que

$$\begin{aligned}
\alpha &= \frac{\frac{R^2 \theta_2^2 (\theta_2 + R)^4}{2(2\theta_2 + R)^4} \left[(2\theta_2 + R)^2 + 16\theta_2^2 \right]}{\left(\frac{1}{4} \right) \left(\frac{R^6 \theta_2^2 (\theta_2 + R)^2}{(2\theta_2 + R)^4} \right)} \\
&= \frac{2 (\theta_2 + R)^2 \left[(2\theta_2 + R)^2 + 16\theta_2^2 \right]}{R^4}
\end{aligned}$$

$$\begin{aligned}\beta &= \frac{\frac{R^2 \theta_2^3 (\theta_2 + R)^5}{(2\theta_2 + R)^4} (10\theta_2 + R)}{\left(\frac{1}{4}\right) \left(\frac{R^6 \theta_2^2 (\theta_2 + R)^2}{(2\theta_2 + R)^4}\right)} \\ &= \frac{4\theta_2 (\theta_2 + R)^3 (10\theta_2 + R)}{R^4}\end{aligned}$$

$$\begin{aligned}\gamma &= \frac{\frac{10R^2 \theta_2^4 (\theta_2 + R)^6}{(2\theta_2 + R)^4}}{\left(\frac{1}{4}\right) \left(\frac{R^6 \theta_2^2 (\theta_2 + R)^2}{(2\theta_2 + R)^4}\right)} \\ &= \frac{40\theta_2^2 (\theta_2 + R)^4}{R^4}\end{aligned}$$

Por lo tanto

$$\begin{aligned}M^{-1}(\xi^*) &= \begin{pmatrix} \frac{2(\theta_2 + R)^2 [(2\theta_2 + R)^2 + 16\theta_2^2]}{R^4} & \frac{4\theta_2 (\theta_2 + R)^3 (10\theta_2 + R)}{\theta_1 R^4} \\ \frac{4\theta_2 (\theta_2 + R)^3 (10\theta_2 + R)}{\theta_1 R^4} & \frac{20\theta_2^2 (\theta_2 + R)^4}{\theta_1^2 R^4} \end{pmatrix} \\ &= \frac{2(\theta_2 + R)^2}{R^4} \begin{pmatrix} (2\theta_2 + R)^2 + 16\theta_2^2 & \frac{2\theta_2 (\theta_2 + R)(10\theta_2 + R)}{\theta_1} \\ \frac{2\theta_2 (\theta_2 + R)(10\theta_2 + R)}{\theta_1} & \frac{20\theta_2^2 (\theta_2 + R)^2}{\theta_1^2} \end{pmatrix}\end{aligned}$$

Una vez obtenida la inversa de la matriz de información evaluada en el diseño óptimo, se calcula la función de varianza

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

donde $f(x) = \left(\frac{x}{\theta_2+x}, -\frac{\theta_1 x}{(\theta_2+x)^2} \right)^T$,

$$\begin{aligned}
d(x, \xi) &= \begin{pmatrix} \frac{x}{\theta_2+x} \\ -\frac{\theta_1 x}{(\theta_2+x)^2} \end{pmatrix}^T \frac{2(\theta_2+R)^2}{R^4} \begin{pmatrix} (2\theta_2+R)^2 + 16\theta_2^2 & \frac{2\theta_2(\theta_2+R)(10\theta_2+R)}{\theta_1} \\ \frac{2\theta_2(\theta_2+R)(10\theta_2+R)}{\theta_1} & \frac{20\theta_2^2(\theta_2+R)^2}{\theta_1^2} \end{pmatrix} \begin{pmatrix} \frac{x}{\theta_2+x} \\ -\frac{\theta_1 x}{(\theta_2+x)^2} \end{pmatrix} \\
&= \frac{2(\theta_2+R)^2}{R^4} \begin{pmatrix} \frac{x(2\theta_2+R)^2(\theta_2+x) + 16\theta_2^2 x(\theta_2+x) - 2\theta_2 x(\theta_2+R)(10\theta_2+R)}{(\theta_2+x)^2} \\ \frac{2\theta_2 x(\theta_2+R)(\theta_2+x)(10\theta_2+R) - 20\theta_2^2 x(\theta_2+R)^2}{\theta_1(\theta_2+x)^2} \end{pmatrix}^T \begin{pmatrix} \frac{x}{\theta_2+x} \\ -\frac{\theta_1 x}{(\theta_2+x)^2} \end{pmatrix} \\
&= \frac{2(\theta_2+R)^2}{R^4} \left(\frac{x^2(2\theta_2+R)^2(\theta_2+x) + 16\theta_2^2 x^2(\theta_2+x) - 2\theta_2 x^2(\theta_2+R)(10\theta_2+R)}{(\theta_2+x)^3} \right) \\
&\quad - \frac{2(\theta_2+R)^2}{R^4} \left(\frac{2\theta_2 x^2(\theta_2+R)(\theta_2+x)(10\theta_2+R) - 20\theta_2^2 x^2(\theta_2+R)^2}{(\theta_2+x)^4} \right) \\
&\quad - \frac{2(\theta_2+R)^2}{R^4} (x^2(2\theta_2+R)^2(\theta_2+x) + 16\theta_2^2 x^2(\theta_2+x) - 4\theta_2 x^2(\theta_2+R)(\theta_2+x)(10\theta_2+R)) \\
&\quad + \frac{2(\theta_2+R)^2}{R^4} (20\theta_2^2 x^2(\theta_2+R)^2)
\end{aligned}$$

cuyo máximo, 2, se alcanza en $x = \frac{R\theta_2}{2\theta_2+R}$ y $x = R$. Puesto que 2 es el número de parámetros, el teorema de equivalencia nos dice que ξ^* es un diseño G-óptimo, y por tanto D-óptimo.

Ahora, consideremos que la concentración x se encuentra entre 0 y 1, y utilizando la estimación de los parámetros en el capítulo 3, en donde $\theta^T = (212.7, 0.0641)^T$, se tiene que el diseño óptimo es

$$\xi^* = \begin{pmatrix} 0.0568 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

En la Figura (6.1) se muestra que la función, $d(x; \xi^*)$ tiene todos sus valores por debajo de cero, y en los puntos de soporte alcanza su máximo, luego ξ^* es D-óptimo para estimar el vector de parámetros θ .

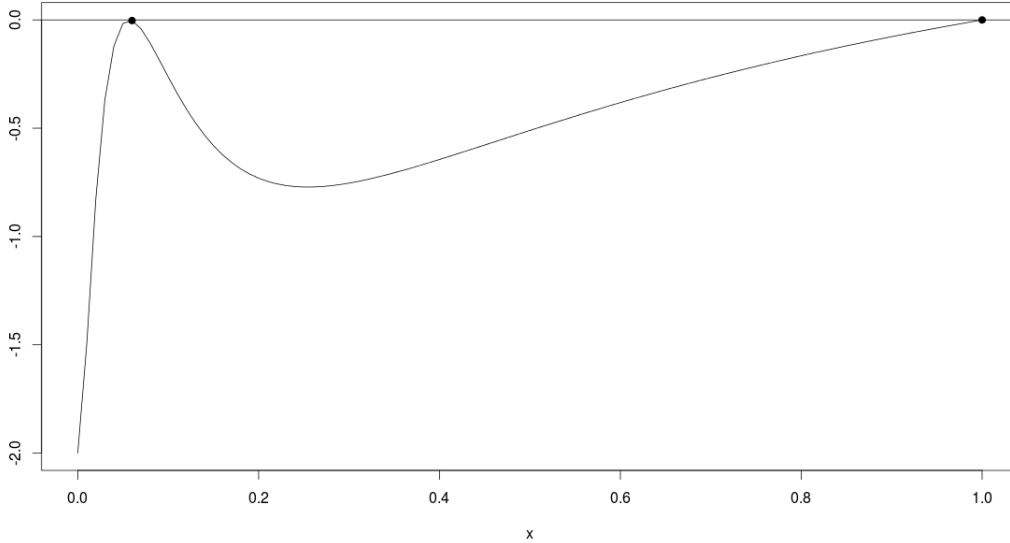


Figura 6.1: Gráfico de la función de varianza $d(x, \xi^*)$, del modelo de Michaelis-Menten.

6.5. Diseños D-óptimos Bayesianos.

Enfoques bayesianos para el diseño óptimo de experimentos en el modelo estadístico no lineal han sido revisados recientemente por Atkinson et al. (2007) y por Chaloner y Verdinelli (1995). Para el modelo lineal de la teoría normal con varianza conocida σ^2 , supongamos que la información anterior es que θ se distribuye normalmente con media θ_0 y matriz de covarianza $\sigma^2 M^{-1}(n_0)$. La información de la matriz $M(n_0)$ formaliza la información previa como equivalentes a los resultados de las observaciones, donde n_0 no necesita ser un número entero. El diseño de las observaciones anteriores se puede escribir $n_0 M(\xi_0)$, cuando $M(n_0) = n_0 M(\xi_0)$. El experimento consiste en n ensayos con diseño de medida ξ_n . La matriz de información para la distribución posterior de θ es

$$\tilde{M}(\xi_n) = n_0 M(\xi_0) + n M(\xi_n). \quad (6.11)$$

Diseños D-óptimos bayesianos para el modelo lineal maximizan $|\tilde{M}(\xi)|$, que hace que no depende del valor apriori de θ_0 . Estos diseños maximizan la ganancia esperada en la información de Shannon entre

las distribuciones aprioris y posteriores de θ :

$$\int \log \frac{p(\theta|y, \xi_n)}{p(\theta)} p(\theta|y, \xi_n) d\theta dy. \quad (6.12)$$

La distribución apriori no depende del diseño ξ , por lo que el diseño ξ maximiza la ganancia esperada en la información de Shannon es el que maximiza

$$U(\xi_n) = \int \log \{p(\theta|y, \xi_n)\} p(\theta|y, \xi_n) d\theta dy. \quad (6.13)$$

Esta es la información esperado de Shannon de la distribución posteriori. Esta utilidad esperada $U(\xi)$ podría ser apropiada cuando el experimento se lleva a cabo para la inferencia en el vector de β . La información apriori se vuelve menos importante, es que $n_0/n \rightarrow 0$, el criterio bayesiano (6.11) tiende a D-optimalidad estándar. La ecuación (6.11) es la misma que la utilizada en el aumento de diseños frequentistas, cuando los ensayos N_0 forman el diseño de un experimento anterior y los resultados de todos los $n_0 + n$ ensayos son para ser analizados en conjunto.

La teoría de los modelos no lineales $\eta(x_i, \theta)$, el enfoque bayesiano para encontrar el diseño óptimo exacto consiste en especificar la distribución apriori de θ que es $p(\theta)$, es más complicado y requiere el uso de aproximaciones a la distribución del estimador $\tilde{\theta}$ posteriori. Si la matriz de información previa en el equivalente de (6.11) se ignora y el estimador de mínimos cuadrados $\hat{\theta}$ se utiliza en lugar de $\tilde{\theta}$, el estimador será asintóticamente normal con una distribución que sólo depende de la matriz de información del experimento. Luego, aproximadamente, el diseño maximiza la información esperada de Shannon que maximizará la utilidad esperada

$$\begin{aligned} \Phi(\xi_n) &= E_{\theta} \log |M(\theta, \xi_n)| \\ &= \int_{\theta} \log |M(\theta, \xi_n)| p(\theta) d\theta. \end{aligned} \quad (6.14)$$

donde $M(\theta, \xi_n)$ es la matriz de información esperada de Fisher definida por la ecuación (6.5) y (6.14) se conoce como criterio D-óptimo bayesiano. Tenga en cuenta que esta enfoque bayesiano no implica que los métodos bayesianos de análisis de los datos se van a utilizar (Murphy et al., 2003). Este enfoque es de vez en cuando, y tal vez más apropiadamente, llamado diseño óptimo promedio, pero el nombre bayesiano se ha convertido en norma y vamos a seguir para utilizarla. Después de haber obtenido una expresión para el criterio (6.14), tenemos que evaluar su esperanza sobre la distribución apriori de θ

para cada diseño considerado. Dado que la integral es casi siempre imposible de evaluar, nosotros lo aproximamos numéricamente mediante el muestreo de $p(\theta)$. Supongamos que probamos m puntos de distribución de cada parámetro para obtener $\theta_1, \dots, \theta_m$ y, para cada diseño considerado, aproximada $\Phi(\xi_n)$ por

$$\tilde{\Phi}(\xi_n) = \frac{1}{m} \left\{ \sum_{i=1}^m \log |M(\theta = \tilde{\theta}_j, \xi_n)| \right\}. \quad (6.15)$$

Para el caso en que no se tiene información sobre los parámetros se utilizaron distribuciones uniformes, es decir, $\theta_1 \sim U(a, b)$ y $\theta_2 \sim U(c, d)$, a continuación se muestra el proceso de maximización de la información de Shannon con dichas distribuciones.

$$\begin{aligned} E_\theta [\log |M(\xi, \theta)|] &= \int_\theta \log |M(\xi, \theta)| d\pi(\theta) \\ &= \int_c^d \int_a^b [2 \log \theta_1 + 2 \log x_1 + \log w_1 + 2 \log(x_1 - x_2) + 2 \log x_2] \frac{d\theta_1 d\theta_2}{(b-a)(d-c)} \\ &\quad + \int_c^d \int_a^b [\log(1 - w_1) - 4 \log(\theta_2 + x_1) - 4 \log(\theta_2 + x_2)] \frac{d\theta_1 d\theta_2}{(b-a)(d-c)} \\ &= \int_c^d \int_a^b \frac{2 \log \theta_1 d\theta_1 d\theta_2}{(b-a)(d-c)} + \int_c^d \int_a^b \frac{2 \log x_1 d\theta_1 d\theta_2}{(b-a)(d-c)} + \int_c^d \int_a^b \frac{\log w_1 d\theta_1 d\theta_2}{(b-a)(d-c)} \\ &\quad + \int_c^d \int_a^b \frac{2 \log(x_1 - x_2) d\theta_1 d\theta_2}{(b-a)(d-c)} + \int_c^d \int_a^b \frac{2 \log x_2 d\theta_1 d\theta_2}{(b-a)(d-c)} + \int_c^d \int_a^b \frac{\log(1 - w_1) d\theta_1 d\theta_2}{(b-a)(d-c)} \\ &\quad - \int_c^d \int_a^b \frac{4 \log(\theta_2 + x_1) d\theta_1 d\theta_2}{(b-a)(d-c)} - \int_c^d \int_a^b \frac{4 \log(\theta_2 + x_2) d\theta_1 d\theta_2}{(b-a)(d-c)} \\ &= \int_c^d \frac{2(b \log b - b - a \log a + a) d\theta_2}{(b-a)(d-c)} + \int_c^d \frac{2 \log x_1 d\theta_2}{(d-c)} + \int_c^d \frac{\log w_1 d\theta_2}{(d-c)} \\ &\quad + \int_c^d \frac{2 \log(x_1 - x_2) d\theta_2}{(d-c)} + \int_c^d \frac{2 \log x_2 d\theta_2}{(d-c)} + \int_c^d \frac{\log(1 - w_1) d\theta_2}{(d-c)} \\ &\quad - \int_c^d \frac{4 \log(\theta_2 + x_1) d\theta_2}{(d-c)} - \int_c^d \frac{4 \log(\theta_2 + x_2) d\theta_2}{(d-c)} \\ &= \frac{2(b \log b - b - a \log a + a)}{(b-a)} + 2 \log x_1 + \log w_1 + 2 \log(x_1 - x_2) + 2 \log x_2 + \log(1 - w_1) \\ &\quad - 4 \left[\frac{(d + x_1) \log(d + x_1) - (d + x_1) - (c + x_1) \log(c + x_1) + (c + x_1)}{d - c} \right] \\ &\quad - 4 \left[\frac{(d + x_2) \log(d + x_2) - (d + x_2) - (c + x_2) \log(c + x_2) + (c + x_2)}{d - c} \right]. \quad (6.16) \end{aligned}$$

Si $x_1 < x_2$ esta función es creciente en x_2 en $[L, R]$, así $x_2 = R$. Por lo que el problema se reduce en

maximizar con respecto a x_1 , w_1 y w_2 . Derivando con respecto a x_1 , y w_1 6.16 se tiene

$$\frac{\partial E_\theta [\log |M(\xi, \theta)|]}{\partial x_1} = \frac{2}{x_1} + \frac{2}{x_1 - R} - 4 \left[\frac{\log(d + x_1) - \log(c + x_1)}{d - c} \right], \quad (6.17)$$

$$\begin{aligned} \frac{-\partial \log |M(\xi)|}{\partial w_1} &= -\frac{1}{w_1} + \frac{1}{1 - w_1} \\ &= \frac{-1 + w_1 + w_1}{w_1(1 - w_1)} \\ &= \frac{1 - 2w_1}{w_1(1 - w_1)} \end{aligned} \quad (6.18)$$

Igualando a cero y despejando x_1 y w_1 de (6.17) y (6.18) respectivamente, se obtiene los valores deseados. Para encontrar el valor de x_1 se realizó numericamente ya que 6.17 no tiene una solución analítica, para ello se utilizó que $\theta_2 \sim U(0, 0.1)$ y $R = 1$. Así el diseño óptimo bayesiano es

$$\xi^* = \begin{pmatrix} 0.0364 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Para el caso en que $\theta = (\theta_1, \theta_2)$ tenga otras distribuciones, el criterio a utilizar es la aproximación a la información de Shannon (6.15).

Capítulo 7

Conclusiones.

En esta tesis se describe el diseño óptimo de experimentos no lineal del modelo de Michaelis-Menten, con aplicaciones a ejemplos de la vida real que se plantean en la cinética enzimática. Aunque la aplicación de estas técnicas ha sido la novela en sí misma, esta tesis ha presentado los métodos de diseño de experimentos con el modelo no lineal.

En primer lugar nos centramos en lo que es la cinética de enzimas, la relación entre la cinética enzimática y el modelo de Michaelis-Menten y luego se discute el mecanismo del modelo de Michaelis-Menten en el capítulo 2. En el siguiente capítulo, Capítulo 3, se describe el modelo lineal y el modelo no lineal se describir el procedimiento de estimación de los parámetros. Presentamos una manera mucho más fácil de estimar los parámetros de las estimaciones encontradas utilizando el modelo lineal.

En el capítulo 4 se representa la teoría del diseño óptimo, diversos criterios sobre el diseño óptimo.

En el capítulo 5 se hace una extensa revisión de la teoría sobre el teorema de equivalencia, así como los puntos de soporte y los pesos del diseño.

En el capítulo 6 se discute el diseño óptimo para el modelo de Michaelis-Menten y obtenemos un diseño óptimo de 2 puntos sobre la base del criterio D-óptimo maximizando el determinante de la matriz de información de Fisher, en un punto situado en $R\theta_2 / (2\theta_2 + R)$ en donde se involucra un parámetro de Michaelis-Menten y otro punto situado en el extremo derecho del intervalo donde se desea encontrar el diseño óptimo.

Los modelos óptimos para modelos no lineales necesariamente requieren de un conocimiento previo de los parámetros. Las alternativas que se han estudiado en esta tesis son: la clásica en donde se minimiza un criterio adecuado suponiendo valores puntuales (locales) de los parámetros; la segunda

alternativa postula que la incertidumbre que tenemos sobre los parámetros la representemos mediante una distribución apriori, esto da lugar a los diseños óptimos bayesianos que fueron bien estudiados en el capítulo 6.

Apendice.

Proposición 10:

(1) Directamente de la definición estandar de un determinante obtenemos

$$\det M = \sum_{\pi} (-1)^{\pi} \{M\}_{1\pi(1)} \{M\}_{2\pi(2)} \cdots \{M\}_{m\pi(m)},$$

donde la suma es bajo todas las permutaciones π de m -tuplas $(1, 2, \dots, m)$, y donde $(-1)^{\pi} = 1$ (o $(-1)^{\pi} = -1$) si la permutación $\pi(1), \dots, \pi(m)$ es obtenida por un número par (o un impar) de intercambios de parejas en la m -tupla $(1, 2, \dots, m)$. Se deduce que el $\det M$ es un polinomio homogéneo de elementos de la matriz M . Como una consecuencia la función $M \mapsto -\log \det M$ es continua.

(2) Tenemos que probar que para cada $\beta \in [0, 1]$, $M_1, M_2 \in \mathcal{M}$

$$\log \det [(1 - \beta) M_1 + \beta M_2] \geq (1 - \beta) \log \det M_1 + \beta \log \det M_2.$$

Esta desigualdad se hace evidente cuando $\det M_1 = 0$ o $\det M_2 = 0$. Por lo tanto, suponer que M_1, M_2 son no singulares. Por la descomposición espectral existe una matriz U no singular tal que $UM_1U^T = I$, $UM_2U^T = \Lambda \equiv \text{diag}(\lambda_1, \dots, \lambda_m)$. Usando la estricta concavidad del logaritmo obtenemos

$$\begin{aligned} \log \det [(1 - \beta) M_1 + \beta M_2] &= \log \det U^{-1} [(1 - \beta) I + \beta \Lambda] U^{-1} \\ &\geq \log \det U^{-2} + \sum_{i=1}^m [(1 - \beta) 1 + \beta \lambda_i] \\ &= (1 - \beta) \log \det U^{-2} + \beta \log \det U^{-1} \Lambda U^{-1} \\ &= (1 - \beta) \log \det M_1 + \beta \log \det M_2. \end{aligned}$$

La desigualdad es estricta si $\beta \in (0, 1)$ y $\lambda_i \neq 1$ al menos para un $i \in \{1, \dots, m\}$, es decir, $M_1 \neq M_2$.

(3) Cuando $-\log \det M < \infty$, M es no singular. El conjunto

$$\{A : A \in \mathbb{R}^{m \times m}, |\log \det A - \log \det M| < 1\}$$

es una vecindad de M que contiene matrices no singulares por la continuidad de $\log \det$. Por lo tanto la siguiente derivada es correcta:

$$\begin{aligned} \frac{\partial}{\partial \{M\}_{ij}} [\log \det M] &= \frac{1}{\det M} \frac{\partial \det M}{\partial \{M\}_{ij}} \\ &= \frac{1}{\det M} \frac{\partial}{\partial \{M\}_{ij}} \left[\sum_{k=1}^m \{M\}_{ik} \det M^{(ik)} \right] \\ &= \frac{\det M^{(ij)}}{\det M} \\ &= \{M^{-1}\}_{ij}, \end{aligned}$$

donde $\det M^{(ij)}$ es el complemento algebraico de $\{M\}_{ik}$ en la matriz M .

Proposición 13:

(1) Si $\sum_{i=1}^m \text{var}_{\xi} \alpha_i < \infty$, entonces $\text{var}_{\xi} \alpha_i < \infty$ para $i = 1, \dots, m$, y la continuidad en el punto $M(\xi)$ se sigue de la proposición III.14 (Pazman 1986).

En caso de que $\sum_{i=1}^m \text{var}_{\xi} \alpha_i = \infty$ considere una secuencia $\{M(\xi_n)\}_{n=1}^{\infty}$ tendiendo a $M(\xi)$. Hay al menos un $i \in \{1, \dots, m\}$ tal que $\text{var}_{\xi} \alpha_i = \infty$ así, por la proposición III.14 (Pazman 1986)

$$\lim_{n \rightarrow \infty} \text{var}_{\xi} \alpha_i = \infty.$$

Se sigue que

$$\lim_{n \rightarrow \infty} \inf \sum_{i=1}^m \text{var}_{\xi_n} \alpha_i \geq \lim_{n \rightarrow \infty} \text{var}_{\xi} \alpha_i = \infty.$$

(2) La convexidad es una consecuencia directa de la proposición III.13 (Pazman 1986).

(3) Si $\sum_{i=1}^m \text{var}_{\xi} \alpha_i < \infty$, entonces los parámetros $\alpha_1, \dots, \alpha_m$ son estimables así la matriz $M(\xi)$ es no singular, y

$$\sum_{i=1}^m \text{var}_{\xi} \alpha_i = \text{tr} M^{-1}(\xi).$$

tomemos las derivadas de

$$I = MM^{-1}$$

con respecto a los elementos de la matriz M . Obtenemos

$$0 = \frac{M}{\partial \{M\}_{ij}} M^{-1} + M \frac{\partial M^{-1}}{\partial \{M\}_{ij}}; \quad i, j = 1, \dots, m$$

y

$$\frac{\partial M^{-1}}{\partial \{M\}_{ij}} = -M^{-1} \frac{\partial M}{\partial \{M\}_{ij}} M^{-1}, \quad i, j = 1, \dots, m.$$

Por lo tanto

$$\nabla_M \text{tr} M^{-1} = -M^{-2}.$$

Proposición 15:

(1) Si $\sup_{x \in X} \text{var}_\xi g_x = \infty$, entonces hay un $x_0 \in \mathcal{X}$ tal que $\text{var}_\xi g_{x_0} = \infty$. En efecto, asumiendo que $\text{var}_\xi g_x < \infty$ para cada $x \in \mathcal{X}$, obtenemos que para cada funcional g_x es estimable. Se deduce que $f(x) \in \mathbb{M}[M(\xi)]$; ($x \in \mathcal{X}$, $\mathbb{M}(M) = \{Mu : u \in \mathbb{R}^m\}$) por lo tanto $\mathbb{M}[M(\xi)] = \mathbb{R}^m$, y $M(\xi)$ es no singular. Como una consecuencia, la función

$$x \in \mathcal{X} \mapsto f^T(x) M^{-1} f(x) = \text{var}_\xi g_x$$

es continua, por lo tanto acotado en el conjunto compacto \mathcal{X} .

Supongamos ahora que $\text{var}_\xi g_x = \infty$ y que

$$\lim_{n \rightarrow \infty} M(\xi_n) = M(\xi).$$

de la proposición III.14 (Pazman 1986) se sigue que

$$\lim_{n \rightarrow \infty} \text{var}_{\xi_n} g_{x_0} = \infty,$$

por lo tanto

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathcal{X}} \text{var}_{\xi_n} g_x = \infty = \sup_{x \in \mathcal{X}} \text{var}_\xi g_x.$$

Cuando $\det M_0 \neq 0$, el conjunto $\mathcal{M}_0 \equiv \{M, M_0 : M, M_0 \in \mathcal{M}, \frac{1}{2} \det M_0 < \det M < 2 \det M_0\}$ es un conjunto abierto de \mathcal{M}_+ . Por lo tanto

$$\phi(M) = \max_{x \in \mathcal{X}} f^T(x) M^{-1} f(x); \quad M \in \mathcal{M}_0,$$

así la continuidad de Φ_G en \mathcal{M}_+ es una consecuencia de la proposición IV.4 (Pazman 1986).

(2) La convexidad de Φ_G se sigue de la proposición III.13 (Pazman 1986). Denotemos por x_0 al punto máximo de la función

$$x \in \mathcal{X} \mapsto f^T(x) [(1 - \beta) M_1 + \beta M_2]^{-1} f(x).$$

de la proposición III.13 se sigue que

$$\begin{aligned} f^T(x_0) [(1 - \beta) M_1 + \beta M_2]^{-1} f(x_0) &< (1 - \beta) f^T(x_0) M^{-1} f(x_0) + \beta f^T(x_0) M^{-1} f(x_0) \\ &\leq (1 - \beta) \max_{x \in \mathcal{X}} f^T(x) M_1^{-1} f(x) + \beta \max_{x \in \mathcal{X}} f^T(x) M_2^{-1} f(x), \end{aligned}$$

es decir, la convexidad estricta de Φ_G .

Proposición 17:

(1) supongamos que

$$\lim_{n \rightarrow \infty} M(\xi_n) = M(\xi).$$

Si $\Phi_E[M(\xi)] = \infty$, entonces hay un vector $g \notin \mathbb{M}[M(\xi)]$; por lo tanto, según la proposición III.14 (Pazman 1986), $\lim_{n \rightarrow \infty} \text{var}_{\xi_n} g = \text{var}_{\xi} g = \infty$. De

$$\Phi_E[M(\xi_n)] \geq \text{var}_{\xi_n} g, \quad n = 1, 2, \dots$$

se sigue que

$$\lim_{n \rightarrow \infty} \Phi_E[M(\xi_n)] = \infty.$$

Si $\det M(\xi) \neq 0$ podemos escribir

$$\Phi_E(M) = \max \{g^T M^{-1} g : g \in G_1\}$$

para cada M perteneciente al conjunto

$$\mathcal{M}_0 \equiv \left\{ M : M \in \mathcal{M}, \frac{1}{2} \det M(\xi) < \det M < 2 \det M(\xi) \right\}.$$

(Observe que G_1 es un conjunto compacto, por lo tanto $\sup = \max$) La continuidad de Φ_E en \mathcal{M}_0 se sigue de proposición IV.4 (Pazman 1986). Por lo tanto \mathcal{M}_0 es abierto en \mathcal{M} , la función Φ_E es continua en \mathcal{M} en el punto M .

(2) Sea $M_1, M_2 \in \mathcal{M}$, $\det M_1 \neq 0$, $\det M_2 \neq 0$, $\beta \in (0, 1)$. Denotemos por g_0 al vector correspondiente

al máximo de la función

$$g \in G_1 \mapsto g^T [(1 - \beta) M_1 + \beta M_2]^{-1} g.$$

Se sigue de la proposición III.13 que

$$\begin{aligned} \Phi_E [(1 - \beta) M_1 + \beta M_2] &= g_0^T [(1 - \beta) M_1 + \beta M_2]^{-1} g_0 \\ &< (1 - \beta) g_0^T M_1^{-1} g_0 + \beta g_0^T M_2^{-1} g_0 \\ &\leq (1 - \beta) \Phi_E (M_1) + \beta \Phi_E (M_2), \end{aligned}$$

que implica la convexidad estricta de Φ_E en el conjunto \mathcal{M}_+ . La convexidad en el conjunto \mathcal{M} se sigue, desde $\det M = 0$ implica que $\Phi_G(M) = \infty$.

Proposición 20:

(1) Por la descomposición espectral existe una matriz no singular U tal que

$$U^T M U = I.$$

Podemos escribir

$$\Phi_E(M) = \begin{cases} \text{tr}(U^T M U)^{-1} & \text{si } \det(U^T M U) \neq 0, \\ \infty & \text{en otro caso.} \end{cases}$$

El mapeo

$$\Psi_1 : M \in \mathcal{M} \mapsto U^T M U$$

es lineal y es continuo. El conjunto $\Psi_1(\mathcal{M}) = \{U^T M U : M \in \mathcal{M}\}$ es el conjunto de toda la información de matrices en el experimento considerado $(\mathcal{X}, \Theta, 1)$ pero con una base lineal cambiada de $\Theta : (U^T f)_1, \dots, (U^T f)_m$. Esto puede ser visto de

$$U^T M(\xi) U = \sum_{x \in \mathcal{X}} [U^T f(x)] [U^T f(x)]^T \xi(x).$$

Definamos otra función Ψ_2 en el conjunto $\Psi_1(\mathcal{M})$ por

$$\Psi_2(U^T M U) = \begin{cases} \text{tr}(U^T M U)^{-1} & \text{si } \det(U^T M U) \neq 0, \\ \infty & \text{si } \det(U^T M U) = 0. \end{cases}$$

De acuerdo a la proposición 13, la función Ψ_2 es continua lo que implica la continuidad de $\Phi_L = \Psi_2 \circ \Psi_1$.

(2) La convexidad requerida de Φ_L se sigue de la convexidad de Ψ_2 (ver proposición 13) y de la estricta convexidad de Ψ_2 en el conjunto $\{\Psi_1(M) : M \in \mathcal{M}, \det M \neq 0\}$.

(3) Usando la proposición 13 obtenemos

$$\begin{aligned} \{\nabla_M \text{tr} W M^{-1}\}_{ij} &= \frac{\partial \Psi_2[\Psi_1(M)]}{\partial \{M\}_{ij}} \\ &= -\{U\}_j. (U^T M U)^{-2} \{U^T\}_{.j}, \end{aligned}$$

que implica lo que queriamos demostrar.

Proposición 23:

(1) El mapeo

$$M \in \mathcal{M}_+ \mapsto H M^{-1} H^T$$

es continua (ver proposición proposición IV.4 (Pazman 1986)). Además, el mapeo

$$A \mapsto \|A\|_p$$

es continuo en el conjunto de todas la matrices positivas semidefinidas, así de la proposición IV.11 (Pazman 1986) se sigue que

$$\| \|A\|_p - \|B\|_p \| \leq \|A - B\|_p.$$

Así Φ_L es continua en \mathcal{M}_+ .

Probar ahora la continuidad de Φ_L en $M(\xi) \in \mathcal{M} - \mathcal{M}_+$. Denotemos por g_i el funcional

$$g_i(\vartheta) \equiv \{H\}_i. \alpha; \quad (\vartheta = \alpha^T f \in \Theta).$$

De la proposición 17 se sigue que

$$\text{var}_\xi g_{i_0} = \infty$$

para algún $i_0 \in \{1, \dots, m\}$. Así de la proposición III.14 (Pazman 1986) obtenemos

$$\lim_{n \rightarrow \infty} \text{var}_{\xi_n} g_{i_0} = \infty \tag{7.1}$$

siempre que $\lim_{n \rightarrow \infty} M(\xi_n) = M(\xi)$. Sea $A = HM^{-1}(\xi_n)H^T$, $B = I$, $r = p$ en la desigualdad

$$\frac{1}{m} \text{tr} AB \leq \left[\frac{1}{m} \text{tr} A^r \right]^{1/r} \left[\frac{1}{m} \text{tr} B^{r/(r-1)} \right]^{(r-1)/r}; \quad r > 1.$$

Obtenemos

$$\begin{aligned} \|HM^{-1}(\xi_n)H^T\|_p &\geq \frac{1}{m} \text{tr} HM^{-1}(\xi_n)H^T \\ &= \frac{1}{m} \sum_{i=1}^m \text{var}_{\xi_n} g_i. \end{aligned}$$

Por lo tanto, de la ecuación (7.1)

$$\lim_{n \rightarrow \infty} \|HM^{-1}(\xi_n)H^T\|_p = \infty = \Phi_{L_p}[M(\xi)].$$

(2) Por la proposición III.13 (Pazman 1986), para cada $\beta \in (0, 1)$, $M_1, M_2 \in \mathcal{M}_+$, $u \in \mathbb{R}^m$, $u \neq 0$ tenemos

$$u^T H [(1 - \beta) M_1 + \beta M_2]^{-1} H^T u < (1 - \beta) u^T H M_1^{-1} H^T u + \beta u^T H M_2^{-1} H^T u.$$

Usando la proposición IV.11 (Pazman 1986) a la hora de establecer las matrices

$$(1 - \beta) H M_1^{-1} H^T + \beta H M_2^{-1} H^T$$

y

$$H [(1 - \beta) M_1 + \beta M_2]^{-1} H^T$$

obtenemos

$$\|H [(1 - \beta) M_1 + \beta M_2]^{-1} H^T\|_p < \|(1 - \beta) H M_1^{-1} H^T + \beta H M_2^{-1} H^T\|_p. \quad (7.2)$$

Sea $A = (1 - \beta) H M_1^{-1} H^T$, $B = \beta H M_2^{-1} H^T$ en $[\frac{1}{m} \text{tr}(A + B)]^{1/r} \leq [\frac{1}{m} \text{tr} A^r]^{1/r} + [\frac{1}{m} \text{tr} B^r]^{1/r}$ y comparando con la ecuación (7.2), obtenemos la estricta convexidad de Φ_{L_p} en \mathcal{M}_+ . La convexidad de \mathcal{M} se sigue del hecho de que $\det M = 0$ implica $\Phi_{L_p}(M) = \infty$.

(3) Denotemos por $I^{(i,j)}$ la matriz de $m \times m$

$$\left\{ I^{(i,j)} \right\}_{k,l} = \begin{cases} 0 & \text{si } i \neq k \text{ o } j \neq l, \\ 1 & \text{si } i = k \text{ o } j = l. \end{cases}$$

Por inducción con respecto a p se puede verificar que

$$\frac{\partial A^{-p}}{\partial \{A\}_{ij}} = - \sum_{h=0}^{p-1} A^{-h-1} I^{(i,j)} A^{-p+h},$$

para cada matriz positiva definida A , así que

$$\nabla_{A \text{tr}} (W A^{-p}) = - \sum_{h=0}^{p-1} A^{-p+h} W A^{-h-1}, \quad (7.3)$$

para cada matriz simétrica W . En particular el caso cuando $W = I$ obtenemos

$$\nabla_{A \text{tr}} A^{-p} = -p A^{-p-1}.$$

Usando la ecuación (7.3) obtenemos lo que queríamos demostrar.

Bibliografía

- [1] Atkins, J. and Cheng, C. (1999). Optimal regression designs in the presence of random block effects. *Journal of statistical planning and inference*, 77(2):321–335.
- [2] Atkinson, A. and Donev, A. (1989). The construction of exact d-optimum experimental designs with application to blocking response surface designs. *Biometrika*, 76(3):515.
- [3] Atkinson, A., Donev, A., and Tobias, R. (2007). Optimum experimental designs, with SAS, volume 34. Oxford University Press, USA.
- [4] Bates, D. and Watts, D. (1988). *Nonlinear regression analysis and its applications*.
- [5] Beal, S. and Sheiner, L. (1980). The nonmem system. *The American Statistician*, 34(2):118–119.
- [6] Beal, S. and Sheiner, L. (1982). Estimating population kinetics. *Critical reviews in biomedical engineering*, 8(3):195.
- [7] Berger, M. and Tan, F. (2004). Robust designs for linear mixed effects models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(4):569–581.
- [8] Bohachevsky, I., Johnson, M., and Stein, M. (1986). Generalized simulated annealing for function optimization. *Technometrics*, pages 209–217.
- [9] Box, G. and Lucas, H. (1959). Design of experiments in non-linear situations. *Biometrika*, 46(1/2):77–90.
- [10] Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304.
- [11] Cheng, C. (1995). Optimal regression designs under random block-effects models. *Statistica Sinica*, 5:485–497.

- [12] Cook, R. and Nachtsheim, C. (1980). A comparison of algorithms for constructing exact d-optimal designs. *Technometrics*, 22(3):315–324.
- [13] Cook, R. and Weisberg, S. (1990). Confidence curves in nonlinear regression. *Journal of the American Statistical Association*, 85(410):544–551.
- [14] D’Argenio, D. (1981). Optimal sampling times for pharmacokinetic experiments. *Journal of Pharmacokinetics and Pharmacodynamics*, 9(6):739–756.
- [15] Davidian, M. and Giltinan, D. (1995). *Nonlinear models for repeated measurement data*, volume 62. Chapman & Hall/CRC.
- [16] Davidian, M. and Giltinan, D. (2003). Nonlinear models for repeated measurement data: an overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(4):387–419.
- [17] Dette, H. and Biedermann, S. (2003). Robust and efficient designs for the michaelis-menten model. *Journal of the American Statistical Association*, 98(463):679–686.
- [18] Dette, H. and Haines, L. (1994). E-optimal designs for linear and nonlinear models with two parameters. *Biometrika*, 81(4):739.
- [19] Dette, H., Melas, V., and Pepelyshev, A. (2003). Standardized maximin e-optimal designs for the michaelis-menten model. *Statistica Sinica*, 13(4):1147–1164.
- [20] Dette, H., Melas, V., and Pepelyshev, A. (2004). Optimal designs for a class of nonlinear regression models. *The Annals of Statistics*, 32(5):2142–2167.
- [21] Dette, H., Melas, V., and Wong, W. (2005). Optimal design for goodness-of-fit of the michaelis-menten enzyme kinetic function. *Journal of the American Statistical Association*, 100(472):1370–1381.
- [22] Dette, H. and Neugebauer, H. (1997). Bayesian d-optimal designs for exponential regression models. *Journal of Statistical Planning and Inference*, 60(2):331–349.
- [23] Dette, H. and O’Brien, T. (1999). Optimality criteria for regression models based on predicted variance. *Biometrika*, 86(1):93.

- [24] Dette, H. and Sahn, M. (1998). Minimax optimal designs in nonlinear regression models. *Statistica Sinica*, 8:1249–1264.
- [25] Federov, V. (1972). Theory of optimal experiments.
- [26] Ford, I., Titterton, D., and Kitsos, C. (1989). Recent advances in nonlinear experimental design. *Technometrics*, 31(1):49–60.
- [27] Goos, P. (2002). The optimal design of blocked and split-plot experiments, volume 164. Springer Verlag.
- [28] Haines, L. (1993). Optimal design for nonlinear regression models. *Communications in Statistics-Theory and Methods*, 22(6):1613–1627.
- [29] Han, C. and Chaloner, K. (2003). D-and c-optimal designs for exponential regression models used in viral dynamics and other applications. *Journal of Statistical Planning and inference*, 115(2):585–601.
- [30] Hedayat, A., Zhong, J., and Nie, L. (2003). 2-parameter nonlinear models. Optimal and efficient designs for *Journal of statistical planning and inference*, 124(1):205–217.
- [31] Johnson, M. and Nachtsheim, C. (1983). Some guidelines for constructing exact d- optimal designs on convex design spaces. *Technometrics*, 25(3):271–277.
- [32] Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 30:271–294.
- [33] Lehninger, A., Nelson, D., and Cox, M. (2005). *Lehninger principles of biochemistry*, volume 1. Wh Freeman.
- [34] Lindstrom, M. and Bates, D. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3):673–687.
- [35] Matthews, J. and Allcock, G. (2004). Optimal designs for michaelis–menten kinetic studies. *Statistics in medicine*, 23(3):477–491.
- [36] Mentre, F., Mallet, A., and Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika*, 84(2):429.

- [37] Mitchell, T. (1974). An algorithm for the construction of “d-optimal” experimental designs. *Technometrics*, 16(2):203–210.
- [38] Mitchell, T. and Miller Jr, F. (1970). Use of design repair to construct designs for special linear models. *Math. Div. Ann. Progr. Rept.*(ORNL-4661), pages 130–131.
- [39] Mukhopadhyay, S. and Haines, L. (1995). Bayesian d-optimal designs for the exponential growth model. *Journal of statistical planning and inference*, 44(3):385–397.
- [40] Murphy, E., Gilmour, S., and Crabbe, M. (2003). Efficient and accurate experimental design for enzyme kinetics: Bayesian studies reveal a systematic approach. *Journal of biochemical and biophysical methods*, 55(2):155–178.
- [41] Murray, R., Granner, D., Mayes, P., Rodwell, V., et al. (2000). *Harper’s biochemistry*, volume 599. Appleton & Lange Norwalk (Conn.).
- [42] Pazman, A. *Foundations of optimum experimental design*. (1986). D. Reidel Publishing Company, Dordrecht, The Netherlands.
- [43] Pinheiro, J. and Bates, D. (2009). *Mixed-effects models in S and S-PLUS*. Springer Verlag.
- [44] Pukelsheim, F. (1993). *Optimal design of experiments*. Wiley Online Library.
- [45] Ratkowsky, D. (1990). *Handbook of nonlinear regression models*. M. Dekker.
- [46] Searle, S., Casella, G., McCulloch, C., et al. (1992). *Variance components*. Wiley Online Library.
- [47] Seber, G. and Wild, C. (2003). *Nonlinear regression*, volume 503. LibreDigital.
- [48] Silvey, S. (1980). *Optimal design*. Chapman & Hall.
- [49] Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12(1/2):1–85.
- [50] Song, D. and Wong, W. (1998). Optimal two-point designs for the michaelis-menten model with heteroscedastic errors. *Communications in Statistics-Theory and Methods*, 27(6):1503–1516.
- [51] Van Schalkwyk, D. (1971). On the design of mixture experiments.

- [52] Vonesh, E. (1996). A note on the use of laplace's approximation for nonlinear mixed- effects models. *Biometrika*, 83(2):447.
- [53] Vonesh, E. and Chinchilli, V. (1997). Linear and nonlinear models for the analysis of repeated measurements, volume 1. CRC.
- [54] Wald, A. (1943). On the efficient design of statistical investigations. *The annals of mathematical statistics*, 14(2):134–140.
- [55] Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80(4):791.
- [56] Wynn, H. (1972). Results in the theory and construction of d-optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 133–147.
- [57] Yang, M. and Stufken, J. (2009). Support points of locally optimal design for nonlinear models with two parameters. *The Annals of Statistics*, 37(1):518-541.