

## Efecto de distribuciones a priori en los diseños D-óptimos Bayesianos para un modelo no lineal correlacionado

### Effect of a priori distributions in Bayesian D-optimal designs for a correlated non-linear model

Juan Carlos Mosquera-Benítez <sup>a,\*</sup>  
Víctor Ignacio López-Ríos <sup>b,\*</sup>

Fecha de Recepción: 05.08.18

Fecha de Aceptación: 08.05.19

DOI: <https://doi.org/10.19053/01217488.v10.n2.2019.9504>

#### Resumen

En la práctica pueden surgir complicaciones a la hora de construir diseños óptimos para modelos de regresión no lineales, uno de los grandes problemas se evidencia cuando las observaciones son correlacionadas, debido a que éstas son tomadas de un mismo individuo, objeto o unidad experimental. Al momento de utilizar el criterio de D-optimalidad este depende tanto del vector de parámetros del modelo como de la estructura de correlación supuesta para el término de error. Una forma de evitar esta dependencia es mediante la inclusión de distribuciones a priori en el criterio de D-optimalidad. En este artículo se estudia el efecto que tiene la escogencia de diferentes distribuciones a priori, tales como las distribuciones Uniforme, Gamma y Lognormal en la obtención de los diseños D-óptimos para un modelo no lineal, cuando los errores presentan diferentes estructuras de correlación. Se hallan los diseños al maximizar el criterio de D-optimalidad aproximado por el método de Monte Carlo. Además, se propone una metodología general que permite hallar diseños D-óptimos para cualquier tipo de modelo no lineal en presencia de observaciones correlacionadas. Finalmente, se propone comparar los diseños encontrados mediante el cálculo de las eficiencias tomando como diseño de referencia el obtenido con la distribución a priori Uniforme. Se aplica la metodología establecida en un caso de estudio, y se concluye que los diseños obtenidos dependen tanto de la estructura de correlación como de la distribución a priori considerada.

**Palabras clave:** Diseño D-óptimo, Modelos no lineales, Estructura de correlación, Matriz de Información de Fisher, Distribuciones a priori.

---

a MSc. en Ciencias-Estadística, Escuela de Estadística, Universidad Nacional de Colombia, Medellín-Colombia.  
\*Correo electrónico: [jmosquerab@unal.edu.co](mailto:jmosquerab@unal.edu.co)

b Profesor Asociado, Escuela de Estadística, Universidad Nacional de Colombia, Grupo de Investigación en Estadística.  
\*Correo electrónico: [vilopez@unal.edu.co](mailto:vilopez@unal.edu.co)

**Abstract**

In practice, complications can arise when constructing optimal designs for non-linear regression models. One of the major problems is when the observations are correlated, since they are taken from the same individual, object or experimental unit. When using the D-optimality criterion, it depends both on the parameter vector of the model and on the correlation structure assumed for the error term. One way to avoid this dependence is through the inclusion of a priori distributions in the D-optimality criterion. In this paper we study the effect of the choice of different a priori distributions, such as the Uniform, Gamma and Lognormal distributions in obtaining the D-optimal designs for a non-linear model, when the errors present different correlation structures. The designs are found by maximizing the approximate D-optimality criterion by the Monte Carlo method. In addition, a general methodology is proposed to find D-optimal designs for any type of non-linear model in the presence of correlated observations. Finally, it is proposed to compare the designs found by calculating the efficiencies taking as a reference design the one obtained with the a priori Uniform distribution. The methodology established in a case study is applied, and it is concluded that the designs obtained depend as much on the correlation structure as on the a priori distribution considered.

**Key words:** D-optimal design, non-linear models, correlation structure, Fisher Information Matrix, a priori distributions.

**1. INTRODUCCIÓN**

Cuando se mencionan los diseños óptimos se tiene una variedad de definiciones unas más técnicas que otras, que al final apuntan a lo mismo, pero no se explica de manera clara para que sirven los diseños óptimos, pues bien, un diseño óptimo, puede facilitar el proceso de recolección de datos y posterior análisis de datos, lo que lleva finalmente a conclusiones confiables y reproducibles de una manera rentable a la hora de hacer un experimento.

El gran avance de los diseños óptimos se debe a las importantes contribuciones hechas principalmente por Kiefer (1959) [1] y Kiefer y Wolfowitz (1960) [2] quienes desarrollaron procedimientos computacionales útiles para encontrar diseños óptimos en problemas de regresión, además sugirieron teorías para cuando el interés recae en uno o varios coeficientes de regresión. Posteriormente varios autores como Fedorov (1972) [3], Pukelsheim (1993) [4], entre otros, en sus libros, trataron e hicieron contribuciones en temas relacionados con diseños óptimos de experimentos. Muchos trabajos de diseños, hasta ahora, se centran en modelos lineales, sin embargo, gran variedad de fenómenos naturales que existen se describen a partir de modelos no lineales. La forma general supuesta para dicho modelo no lineal se presenta en la ecuación (1).

$$Y(t_i) = \eta(t_i, \theta) + \varepsilon_i, \quad i = 1, \dots, N \quad (1)$$

donde se supone que la respuesta,  $Y$ , se puede modelar por una función no lineal  $\eta(t_i, \theta)$ , donde  $t_i$  es el valor de la variable explicativa en el  $i$ -ésimo individuo o unidad experimental ( $i=1,2,\dots,N$ ) y  $\varepsilon_i$  es una variable aleatoria con distribución normal con media cero y varianza constante.

Para este tipo de modelos no lineales se destacan los aportes de Box y Lucas (1959) [5] quienes aplicaron el criterio de D-óptimalidad usando un argumento geométrico para la obtención de diseños de  $m$  puntos de un modelo de  $m$  parámetros, que representaba la reacción química de tres sustancias. Años más tarde Box y Hunter (1965) [6] hacen valiosos aportes en los diseños óptimos para modelos no lineales utilizando un método secuencial para la estimación óptima de los parámetros del modelo de tal forma que ésta fuera lo suficientemente precisa.

Usualmente, cuando se construye un diseño óptimo para modelos de regresión no lineales en el vector de parámetros  $\theta$  (ver 1), se supone de antemano que los errores son incorrelacionados, esto conlleva a que cuando se calcula la matriz de información de Fisher asociada al diseño, dependa de este vector de parámetros desconocido, de esta manera la construcción de un diseño óptimo para este proceso depende del vector de parámetros, por lo cual se plantean diferentes enfoques para la construcción de diseños óptimos. A uno de estos enfoques se le denomina diseños óptimos locales introducido por Chernoff (1953) [7], el cual consiste en asignarle un valor a priori  $\theta_0$  muy

cercano al verdadero valor del vector  $\theta$ . Este valor  $\theta_0$  es determinado por el investigador a partir de experiencias previas en experimentos similares o tomando a  $\theta_0$  como un estimador de  $\theta$ . Algunos autores como Ford, Torsney, y Wu (1992) [8], O'Brien (1992) [9], Dette, Melas, y Wong (2005) [10], entre otros, han utilizado este procedimiento para construir diseños óptimos. Otro enfoque para este problema, propuesto por Pronzato y Walter (1985) [11] y Chaloner y Larntz (1989) [12], es utilizar el conocimiento previo que se tiene acerca de  $\theta$  a partir de la escogencia de una distribución a priori  $\pi$ , la cual se incorpora en el criterio de optimalidad, este procedimiento es conocido en la literatura de diseños óptimos como criterio de optimalidad Bayesiano, o también conocido como diseños óptimos promediados por una distribución a priori.

En el caso de los modelos no lineales con observaciones correlacionadas, es decir para el vector  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^T$ , donde  $\varepsilon \sim N_N(0, \Sigma_\beta)$ , la matriz de información asociada a un diseño, además de depender del vector de parámetros  $\theta$ , también depende de otro vector de parámetros  $\beta$  de  $q$  componentes, asociado a la estructura de correlación del término de error y la matriz de información de Fisher. Esta última es una función que depende tanto del diseño como de los dos vectores de parámetros, ver [13]. Para optimizar un funcional de dicha matriz, se escoge el criterio D-optimalidad, el cual consiste en la maximización del determinante de ésta. El diseño D-óptimo depende  $\Theta^T = (\theta^T, \beta^T)$  y es útil en la medida de que tan acertada sea la escogencia de los valores locales. En este tipo de problemas, autores como Ucinski y Atkinson (2004) [14], Amo-Salas, López-Fidalgo, y López-Ríos (2012) [15], Campos-Barreiro y López-Fidalgo (2015) [16], han obtenido diseños óptimos e incluso han implementado nuevos criterios para la obtención de los diseños. Una forma de evitar la dependencia de los valores locales en la construcción de los diseños óptimos es la incorporación de distribuciones a priori en el criterio de optimalidad y por tanto el diseño resultante depende de la escogencia de la distribución a priori. En este artículo se implementa un procedimiento general para hallar diseños óptimos bayesianos para modelos no lineales en presencia de observaciones correlacionadas, para encontrar estos diseños se utiliza el método de Monte Carlo que se

aborda en secciones posteriores, luego el análisis realizado en este artículo se hace considerando el modelo biexponencial que describe la cinética de un fármaco en el organismo y para cada una de la estructuras de correlación propuestas para este modelo se procede a comparar los diseños hallados a través de distribuciones a priori previamente especificadas con un diseño de referencia mediante el cálculo de la eficiencia.

## 2. Diseños óptimos para modelos no lineales

### 2.1 Modelo de regresión y el diseño

Los modelos de regresión no lineal proporcionan una herramienta importante para describir la relación entre una variable respuesta  $Y$  a través de  $k$ -variables explicativas y tienen muchas aplicaciones en ingeniería, física, biología, economía, medicina, entre otras disciplinas [17]. Dicho modelo de regresión no lineal en los parámetros es de la forma:

$$Y(t_i) = \eta(t_i, \theta) + \varepsilon_i, i = 1, \dots, N \quad (2)$$

o en forma matricial el modelo de regresión no lineal se reescribe como:

$$\mathbf{Y} = \eta(\mathbf{t}, \theta) + \varepsilon \quad (3)$$

donde  $\theta \in \mathbb{R}^p$  un vector de  $p$  parámetros desconocidos,  $\mathbf{Y} = (Y(t_1), \dots, Y(t_N))^T$ ,  $\mathbf{t} = (t_1, \dots, t_N)$ ,  $\eta(\mathbf{t}, \theta) = [\eta(t_1, \theta), \dots, \eta(t_N, \theta)]^T$ , donde  $\eta(t_i, \theta)$  es una función de tipo no lineal en  $\theta$ ,  $t_i$  es el valor de la variable explicativa en el  $i$ -ésimo individuo o unidad experimental ( $i=1,2,\dots,N$ ) y toma valores en el espacio del diseño  $\chi \subseteq \mathbb{R}$  y  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^T$  es un vector de errores, donde  $\varepsilon \sim N_N(0, \Sigma_\beta)$ ,  $\beta$  es un vector con  $q$  componentes asociada a la estructura del error.

Un diseño  $\zeta$  se define como una medida de probabilidad con soporte finito sobre el espacio del diseño  $\chi$  y se representa por:

$$\xi = \left\{ \begin{matrix} t_1 & \cdots & t_k \\ w_1 & \cdots & w_k \end{matrix} \right\} \quad (4)$$

donde los  $\{t_1 \dots t_k\}$  son conocidos como  $k$  puntos de soporte distintos del diseño a muestrear, son

elementos del espacio del diseño  $\chi$  con frecuencias (o pesos) aproximadamente proporcionales a  $w_1, w_2, \dots, w_k$  donde  $w_i = \zeta(t_i)$  y  $w_i > 0$  para  $i = 1, \dots, k$ ; tal que  $\sum_i w_i = 1$ . Los  $w_i$  indican la proporción de corridas experimentales que se deben hacer en cada punto del soporte. De la expresión en (4) se desprenden dos tipos de diseños:

- **Diseño Exacto:** El diseño es exacto si los pesos  $w_i = r_i/N$ , con  $r_i$  enteros y representan el número de réplicas a realizarse en  $t_i$  y  $\sum_i r_i = N$ .
- **Diseño aproximado o continuo:** En este caso los pesos  $w_i$  son cualesquier número entre 0 y 1, “es decir”,  $0 < w_i \leq 1$  para  $i = 1, \dots, k$ , por lo cual no se hace ninguna restricción a los pesos de ser números racionales [18].

## 2.2 Estructura de correlación

Considere el modelo de regresión no lineal en (3) expuesto en la sección 2.1, donde los supuestos para este modelo de regresión son los siguientes: Para observaciones correlacionadas  $\varepsilon \sim N_N(0, \Sigma_\beta)$ ,  $\beta$  es un vector con  $q$  componentes asociada a la estructura del error y  $\Sigma_\beta$  está dado por:

$$\Sigma_\beta = [Cov(\varepsilon_i, \varepsilon_j)]_{N \times N} = [E(\varepsilon(t_i)\varepsilon(t_j))]_{N \times N} \quad (5)$$

con

$$Cov(\varepsilon(t_i), \varepsilon(t_j)) = c(|t_i - t_j|, \beta) \quad \forall i, j \quad (6)$$

donde  $c(\cdot, \beta)$  es llamada función de covarianza. Esta función se supone continuamente diferenciable en el espacio parametral [13], el vector  $\beta$  representa los parámetros de la función de covarianza con  $\beta \in \mathbb{R}^q$ .

En [19] reescriben la ecuación (6), con el fin de facilitar los cálculos al utilizar la matriz de información de Fisher.

$$\begin{aligned} Cov(\varepsilon(t_i), \varepsilon(t_j)) &= c(|t_i - t_j|, \beta) \\ &= \sigma^2 c(|t_i - t_j|, \rho) \end{aligned} \quad (7)$$

donde  $\beta^T = (\sigma^2, \rho^T)$  con  $\sigma^2 \in \mathbb{R}$ ,  $\rho \in \mathbb{R}^{q-1}$  parámetros desconocidos en (7).

De esta manera la ecuación (5) se expresa como:

$$\Sigma_\beta = \sigma^2 A_\rho \quad (8)$$

Las estructuras de covarianzas son funciones que tratan de explicar en cierto modo el grado de correlación que existe entre los datos de un determinado experimento. En este trabajo se utilizan las siguientes estructuras de correlación o funciones de covarianza:

Función exponencial

$$Cov(\varepsilon(t_i), \varepsilon(t_j)) = \sigma^2 \exp(-\rho |t_i - t_j|) \quad (9)$$

Función Kernel 1

$$Cov(\varepsilon(t_i), \varepsilon(t_j)) = \sigma^2 (1 + |t_i - t_j|)^{-\rho} \quad (10)$$

Función de Cauchy 1

$$Cov(\varepsilon(t_i), \varepsilon(t_j)) = \sigma^2 (1 + \rho |t_i - t_j|)^{-1/2} \quad (11)$$

La estructura de covarianza en (9) fue tratada por [15] con el fin de estudiar modelos farmacocinéticos anidados, mientras que las estructuras de covarianza en (10) y (11) fueron tratadas por [20] para todo tipo de modelos en presencia de observaciones correlacionadas.

En caso de tener datos, una forma de verificar si los parámetros de la función de covarianza son significativos o no, es mediante el criterio de razón de verosimilitud [21].

## 2.3 Matriz de información

Para cada diseño  $\zeta$  se define la matriz de información de Fisher de acuerdo con las siguientes condiciones:

- Si en (2) los errores tienen la siguiente estructura  $\varepsilon \sim N_N(0, \sigma^2 I)$  la matriz de información de Fisher está dada por:

$$M(\xi, \theta) = \sum_{i=1}^k f(t_i, \theta) f^T(t_i, \theta) w_i \quad (12)$$

donde  $f(t_i, \theta) = \left[ \frac{\partial \eta(t_i, \theta)}{\partial \theta_1}, \dots, \frac{\partial \eta(t_i, \theta)}{\partial \theta_p} \right]^T$ ; para  $i = 1, 2, \dots, k$  en este caso la matriz de información depende del vector de parámetros  $\theta$ .

(b)

**Teorema 2.1.** Si en el modelo en (3) los errores tienen la siguiente estructura  $\varepsilon \sim N_N(0, \Sigma_\beta)$  y  $\xi = \{t_1, t_2, \dots, t_N\}$  es un diseño exacto de  $N$ -puntos, entonces la matriz de información de Fisher está dada por:

$$M(\xi, \theta, \beta) = \begin{pmatrix} M_{11} & 0 \\ 0 & M_{22} \end{pmatrix} \quad (13)$$

donde las matrices  $M_{11}$  y  $M_{22}$  se definen a continuación

$$M_{11} = F^T(\theta) \Sigma_\beta^{-1} F(\theta)$$

$$M_{22} = \left[ \frac{1}{2} \text{tr} \left( \Sigma_\beta^{-1} \frac{\partial \Sigma_\beta}{\partial \beta_k} \Sigma_\beta^{-1} \frac{\partial \Sigma_\beta}{\partial \beta_l} \right) \right]_{q \times q}$$

$$F(\theta) = \left[ \frac{\partial \eta(t_i, \theta)}{\partial \theta_j} \right]_{N \times p}$$

y

$$\Sigma_\beta = [\text{Cov}(\varepsilon_i, \varepsilon_j)]_{N \times N} = [E(\varepsilon(t_i) \varepsilon(t_j))]_{N \times N}$$

La matriz de información de Fisher es una matriz particionada, en donde la diagonal principal de esta matriz está conformada por submatrices cuadradas que dependen de los vectores  $\theta$ ,  $\beta$  y  $\xi$ . La matriz de información en (13) fue expuesta por [13] quien consideró un modelo no lineal donde los errores presentaban una estructura de correlación.

**Corolario 2.1.** Si en el modelo 3,  $\Sigma_\beta = \sigma^2 A_\rho$ , entonces  $M_{11}$  y  $M_{22}$  están dadas por:

$$M_{11} = \frac{1}{\sigma^2} F^T(\theta) A_\rho^{-1} F(\theta)$$

$$M_{22} = \begin{pmatrix} \frac{1}{2\sigma^4} N & \frac{1}{\sigma^2} \text{tr} \left( A_\rho^{-1} \frac{\partial A_\rho}{\partial \rho} \right) \\ \frac{1}{\sigma^2} \text{tr} \left( A_\rho^{-1} \frac{\partial A_\rho}{\partial \rho} \right) & \frac{1}{2} \text{tr} \left( A_\rho^{-1} \frac{\partial A_\rho}{\partial \rho} A_\rho^{-1} \frac{\partial A_\rho}{\partial \rho} \right) \end{pmatrix}$$

## 2.4 Criterio D-optimalidad

El criterio D-optimalidad es un método utilizado para hallar diseños que permite estimar de manera óptima los parámetros del modelo, este criterio se

basa en minimizar el volumen del elipsoide de confianza, asociado a la estimación de máxima verosimilitud de los parámetros del modelo especificado en (2).

El criterio D-optimalidad se define como el siguiente funcional asociado a un diseño:

$$\Psi(\xi; \theta) := \Psi_D(M(\xi; \theta)) = |M^{-1}(\xi; \theta)|^{1/p} \quad (14)$$

donde  $p$ : es el número de parámetros del modelo.

El diseño que resulta de maximizar la expresión en (14) se conoce como diseño D-óptimo.

El criterio D-optimalidad, cuando los errores están correlacionados, se define como el siguiente funcional:

$$\begin{aligned} \Psi(M(\xi; \theta, \beta)) &= |M(\xi; \theta, \beta)|^{1/(p+q)} \\ &= \{ |M_{11}(\xi; \theta)| |M_{22}(\xi; \beta)| \}^{1/(p+q)} \end{aligned} \quad (15)$$

Un diseño  $\xi^*$  es D-óptimo si maximiza (15), esto equivale a tener un diseño que minimice:

$$-\log |M(\xi; \theta, \beta)| = -\log |M_{11}(\xi; \theta)| - \log |M_{22}(\xi; \beta)| \quad (16)$$

Con  $\Psi$  un funcional real y  $|M|$  denota el determinante de la matriz  $M$ , según lo afirmado por [18], es más útil el uso de funciones convexas como el logaritmo natural.

## 3. DISEÑOS ÓPTIMOS BAYESIANOS

En la teoría de los diseños óptimos para modelos no lineales en el vector de parámetros, cuando se calcula la matriz de información de Fisher asociada al diseño  $\xi$ , ésta depende de dos vectores de parámetros, ver (13). Para evitar dicha dependencia se consideran distribuciones a priori  $\pi$  para  $\Theta^T = (\theta^T, \beta^T)$  y se incorpora al criterio de optimalidad, resultando de esta manera un criterio denominado Bayesiano. Un diseño  $\xi$  es  $\Psi$ - óptimo Bayesiano con respecto a la distribución a priori  $\pi$  y se denota como  $\Psi_\pi$  - óptimo, si maximiza a:

$$\begin{aligned} \Psi(\xi, \pi) &= E_\pi \{ \Psi(M(\xi; \theta, \beta)) \} \\ &= \int \Psi(M(\xi; \theta, \beta)) d\pi(\Theta) \end{aligned} \quad (17)$$

En particular, un diseño  $\zeta^*$  es D-óptimo Bayesiano con respecto a una distribución a priori  $\pi$  para el vector de parámetros  $\Theta^T = (\theta^T, \beta^T)$  y se denota como  $D_\pi$ -óptimo si maximiza a:

$$\begin{aligned} \Psi(\xi, \pi) &= E_\pi \{ \log \Psi(M(\xi; \theta, \beta)) \} \\ &= \int \log \left\{ |M(\xi; \theta, \beta)|^{1/(p+q)} \right\} d\pi(\Theta) \end{aligned} \quad (18)$$

En ocasiones, es difícil maximizar la expresión (17), por lo cual se recurre a maximizar el logaritmo de ésta como se muestra en (18).

Se puede mostrar que la expresión dada en (18) se puede reescribir, usando (15) y el Corolario 2.1, para más detalle ver la página 28 de la tesis en [22]:

$$\Psi(\xi, \pi) = \frac{1}{p+q} \int \log \left( |M_{11}(\xi; \theta)| \left| \frac{1}{4\sigma^4} N^2 S_\lambda^2 \right| \right) d\pi(\Theta) \quad (19)$$

donde  $S_\lambda^2 = \frac{1}{N} \sum_{j=1}^N (\lambda_j - \bar{\lambda})^2$ ,  $\lambda_j$  es el  $j$ -ésimo valor propio de la matriz  $\left( A_\rho^{-1} \frac{\partial A_\rho}{\partial \rho} \right)$ .

Para calcular la integral en (18), ésta se calcula usando el método de Monte Carlo que es un método numérico que sirve para calcular integrales que no se pueden resolver analíticamente, este procedimiento se fundamenta en la simulación de variables aleatorias que son generadas del programa estadístico R [23], el método se expone a continuación.

Sea  $N_{sim}$  el número de simulaciones, que corresponde a la cantidad de valores de  $\Theta_i$  que se van a generar de una distribución de probabilidad específica  $\pi(\Theta)$  para cada uno de los parámetros del vector de parámetros  $\Theta$ . Por la ley débil de los grandes números, se tiene que para la distribución  $\pi(\Theta)$  y para la función  $\log|M(\xi; \Theta_i)|^{1/(p+q)}$ .

$$\frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} \log|M(\xi; \Theta_i)|^{1/(p+q)} \xrightarrow{P} \quad (20)$$

$$E_\pi \{ \log|M(\xi; \Theta)|^{1/(p+q)} \}$$

donde la expresión  $X_n \xrightarrow{P} \zeta$  denota que la sucesión de variables aleatorias converge en probabilidad a una constante  $\zeta$  cuando  $n$  tiende a infinito.

Y además por definición de esperanza para distribuciones continuas de probabilidad

$$\begin{aligned} E_\pi \left\{ \log|M(\xi; \Theta)|^{1/(p+q)} \right\} &= \quad (21) \\ \int \log \left\{ |M(\xi; \Theta)|^{1/(p+q)} \right\} d\pi(\Theta) \end{aligned}$$

En otras palabras, de las expresiones dadas en (20) y (21) indican que  $\frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} \log|M(\xi; \Theta_i)|^{1/(p+q)}$  converge en probabilidad al criterio  $D_\pi$ -optimalidad.

Para saber el valor de  $N_{sim}$  apropiado para calcular la integral usando el método de Monte Carlo, se elige una distribución a priori y posteriormente se fijan los valores tanto del diseño como de los parámetros de la integral en (18) y se procede a calcular esta integral vía Monte Carlo utilizando los siguientes valores para  $N_{sim}$ : 1000, 2000, 5000, 7000, 10000, 15000. Y se observa que la integral se estabiliza o se aproxima al valor real cuando  $N_{sim}$  toma un valor de 10000, por tal motivo se decide trabajar con  $N_{sim} = 10000$ .

### 3.1 Eficiencia de un diseño

Una forma de medir la eficiencia de un diseño  $\zeta$ , con respecto a un diseño  $\Psi$ -óptimo  $\zeta^*$ , es utilizando el cociente del criterio de optimalidad evaluado tanto en  $\zeta$  como en  $\zeta^*$ , es decir,

$$\Psi_{ef}(\xi) = \frac{\Psi(\xi)}{\Psi(\xi^*)} \quad (22)$$

donde  $0 < \Psi_{ef}(\xi) \leq 1$ . Valores cercanos a uno de la eficiencia para los diseños  $\zeta^*$   $\Psi$ -óptimo y el diseño  $\zeta$ , significa que el diseño  $\zeta$  es casi equivalente a  $\zeta^*$ , es decir, que los dos diseños tienen casi las mismas propiedades en cuanto a eficiencia. Según [18], la  $D_{ef}(\xi)$  en el caso de D-óptimo local, se puede interpretar como el número de réplicas requeridas del diseño para que sea tan eficiente como el diseño D-óptimo.

Dadas dos distribuciones a priori  $\pi_1$  y  $\pi_2$  para  $\Theta$ , se define la eficiencia del diseño  $\xi$ - $\Psi_{\pi_1}$ -óptimo con respecto al diseño  $\xi^*$ - $\Psi_{\pi_2}$ -óptimo denotado por  $\Psi_{ef \pi_1|\pi_2}(\xi)$  como:

$$\Psi_{ef \pi_1|\pi_2}(\xi) = \frac{E_{\pi_2} \{ \Psi(M(\xi; \theta, \beta)) \}}{E_{\pi_2} \{ \Psi(M(\xi^*; \theta, \beta)) \}} \quad (23)$$

Donde la distribución a priori  $\pi_2$  es la distribución de referencia.

Para el criterio D-optimalidad se define la  $D_{ef \pi_1|\pi_2}(\xi)$  como:

$$D_{ef \pi_1|\pi_2}(\xi) = \frac{\int \log \{ |M(\xi; \theta, \beta)|^{1/(p+q)} \} d\pi_2(\Theta)}{\int \log \{ |M(\xi^*; \theta, \beta)|^{1/(p+q)} \} d\pi_2(\Theta)} \quad (24)$$

La  $D_{ef \pi_1|\pi_2}(\xi)$  suministra la información de que tan eficiente es el diseño  $\xi$  en comparación con el diseño  $\xi^*$ .

### 3.2 Estimación de los parámetros asociados al modelo y a la función de covarianza

En esta sección se da una serie de pasos para estimar los vectores de parámetros asociados tanto al modelo no lineal como a la función de covarianza propuesta.

1. Se elige el conjunto de datos y el modelo no lineal  $Y(t_i) = \eta(t_i, \theta) + \varepsilon_i$ , definido en (2), que describe el comportamiento de dichos datos.
2. Se estima el vector de parámetros  $\theta$  del modelo no lineal utilizando los datos, esto se hace usando la función *nls* del R [23].
3. Si los datos son tomados de un mismo individuo, existe una posible correlación entre ellos, por ende, se proponen funciones de covarianzas de la forma  $Cov(\varepsilon(t_i), \varepsilon(t_j)) = c(|t_i - t_j|, \beta)$ , definida en (6).
4. Se estima el vector de parámetros  $\Theta^T = (\theta^T, \beta^T)$ , conformado por los parámetros tanto del modelo no lineal como de la función de covarianza, esta estimación se hace vía máxima verosimilitud dada por la ecuación (25):

$$l(\Theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_\beta|) - \frac{1}{2} [(\mathbf{Y} - \eta(\mathbf{t}, \theta))^T \Sigma_\beta^{-1} (\mathbf{Y} - \eta(\mathbf{t}, \theta))] \quad (25)$$

mediante la función *nls* del R [23]. Esta función requiere de valores iniciales por lo cual se utiliza  $\Theta_{ini}^T = (\theta_{est}^T, \beta_{tan}^T)$ , donde  $\theta_{est}^T$  es el vector de parámetros estimado en el paso 2 y  $\beta_{tan}^T$  es el vector de parámetros por tanteo, es decir, formado por valores asignados al azar.

Otra forma alternativa de encontrar o de tener una idea de cuáles son los valores del vector de valores iniciales  $\Theta_{ini}^T$ , es mediante la función *DEoptim* del R [23], que es un algoritmo genético de evolución diferencial que sirve para maximizar o minimizar funciones, este algoritmo a la hora de optimizar parte de unos valores aleatorios, y además el algoritmo requiere mucho tiempo para optimizar dichas funciones.

5. Una vez estimado el vector de parámetros  $\beta$ , se efectúa una prueba de hipótesis con la intención de saber si los parámetros de la función de covarianza son significativos o no, en el caso en que no sean significativos quiere decir que la función de covarianza seleccionada no modela la correlación entre los datos por tanto se descarta y se selecciona otra función de covarianza factible.

### 3.3 Metodología para hallar los diseños D-óptimos Bayesianos

Para hallar diseños D-óptimos bayesianos o  $D_\pi$ -óptimos, se utiliza la siguiente metodología

1. Primero se construyen intervalos de confianza bootstrap enunciado en [21] adaptado por los autores para observaciones correlacionadas (ver Apéndice 6.1), para cada uno de los elementos del vector de parámetros estimados  $\hat{\Theta}^T = (\hat{\theta}^T, \hat{\beta}^T)$ .
2. Se considera la distribución uniforme como la primera distribución factible al rango de variación de cada uno de los elementos del vector de parámetros y se toman los extremos del intervalo de confianza bootstrap como el

intervalo donde está definida la distribución uniforme para cada uno de los parámetros del vector  $\hat{\Theta}$ .

3. Con el propósito de comparar los diseños se consideran otras distribuciones a priori factibles al rango de variación de cada uno de los elementos del vector de parámetros  $\Theta$ . Y con el fin de que los diseños sean comparables con los diseños obtenidos con la distribución uniforme se hallan los parámetros de las nuevas a priori igualando la media y la varianza de la distribución uniforme con las respectivas media y varianza de la distribución elegida, este proceso se hace para cada uno de los elementos del vector de parámetros  $\Theta$ .
4. Con las distribuciones a priori para cada parámetro, obtenidas en la parte 3 y bajo independencia, se halla la distribución a priori conjunta, dada en:

$$\pi(\Theta) = \pi_1(\theta_1) \cdot \pi_2(\theta_2) \cdots \pi_p(\theta_p) \pi_{p+1}(\beta_1) \cdot \pi_{p+2}(\beta_2) \cdots \pi_{p+q}(\beta_q) \quad (26)$$

esta a priori conjunta se reemplaza en el criterio D-optimalidad dado en (18).

5. Usando la función *nminb* del R [23], el diseño hallado es  $D_\pi$ -óptimo si maximiza a (18) junto con la a priori conjunta especificada en el paso anterior.
6. Finalmente, para cuantificar el efecto de las distribuciones a priori en los diseños óptimos hallados, se hallan las eficiencias para cada uno de los diseños tomando como referencia a los diseños que se encontraron con la distribución uniforme, estas comparaciones se hacen bajo la misma función de covarianza.

#### 4. CASO DE ESTUDIO: MODELO BIEXPONENCIAL

El siguiente conjunto de datos tomado de [24], representa la concentración de lipoamida (ácido lipoico) en un paciente sano después de administrarle una dosis de 120 mg de este medicamento vía oral antes de desayunar. Se tomaron muestras de sangre en diversos

momentos y se analizaron para detectar el efecto de la lipoamida en el organismo.

**Tabla 1.** Concentración de lipoamida

Tiempo (hora)	Concentración ( $\mu\text{g/L}$ )
0.0	0.0
0.2	38.1
0.4	50.6
0.6	54.1
0.8	54.3
1.0	53.3
1.2	52.0
2.0	46.3
4.0	34.4
10.0	14.1
15.0	6.7
24.0	1.8

En modelo que describe la concentración de lipoamida en distintos momentos del tiempo está dado por:

$$Y(t_i) = \frac{S \cdot F \cdot D \cdot k_a}{V_D(k_a - k_e)} \left[ e^{-k_e t_i} - e^{-k_a t_i} \right] + \varepsilon_i \quad (27)$$

$$\forall i = 1, \dots, 12.$$

donde los parámetros del modelo son  $k_a$  y  $k_e$  son las constantes de absorción y eliminación del medicamento, la dosis inicial es  $D = 120$  mg, el factor de sal se toma como  $S = 1$  y el cociente entre el factor de biodisponibilidad  $F$  y el volumen de distribución del medicamento  $V_D$  está dado por  $F/V_D = 1/2$ , estos valores son determinados previamente por el investigador que está haciendo el estudio, de modo que el modelo en términos de los valores conocidos queda de la siguiente forma:

$$Y(t_i) = \frac{60 \cdot k_a}{(k_a - k_e)} \left[ e^{-k_e t_i} - e^{-k_a t_i} \right] + \varepsilon_i \quad \forall i = 1, \dots, 12. \quad (28)$$

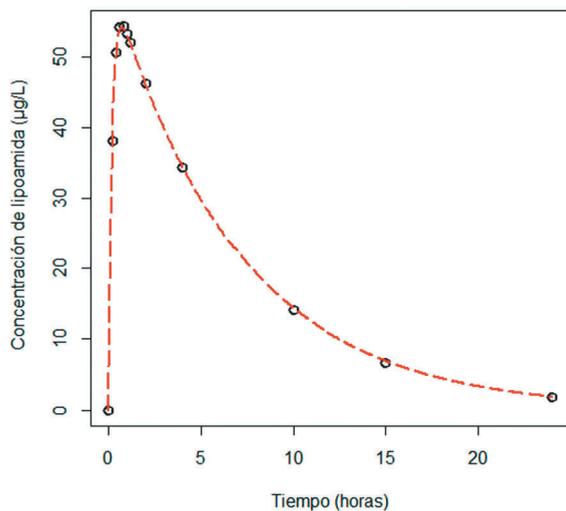
Las estimaciones de los parámetros del modelo no lineal fueron estimados siguiendo los pasos 1 y 2 descritos en la sección 3.2. Los valores se muestran en la Tabla 2.

**Tabla 2.** Estimación de los parámetros del modelo

Parameters	Estimate	Std. Error	Pr(>  t )
$k_a$	5.2197398	0.0434541805	3.922371e-17
$k_e$	0.1459614	0.0009009557	1.972413e-18

De la tabla anterior se observa que la estimación de los parámetros del modelo (28) son altamente significativos, lo que indica que este modelo es el adecuado para modelar la concentración de lipoamida en el cuerpo con respecto al tiempo.

El modelo ajustado para el conjunto de datos que describe la concentración de lipoamida, se muestra en la siguiente figura.



**Figura 1.** Modelo ajustado para concentración de lipoamida

Como los datos son tomados de un mismo individuo, se requiere de una función que exprese el grado de correlación entre estas observaciones, para ello se consideran las funciones de covarianzas expuestas en (9), (10) y (11).

Para ello se estiman los parámetros del modelo y los parámetros de las funciones de covarianzas consideradas usando la función de verosimilitud como se muestra en la siguiente tabla

**Tabla 3.** Estimación de los parámetros del modelo y de la función de covarianza

	Tipo de función de covarianza		
	Exponencial	Kernel 1	Cauchy 1
$\hat{k}_a$	5.1713473	5.17122787	5.17107404
$\hat{k}_e$	0.1468721	0.14691789	0.14688514
$\hat{\sigma}^2$	0.0267699	0.02863071	0.04159506
$\hat{\rho}$	0.5377600	0.74218495	0.80395437

#### 4.1 Diseños para distribuciones a priori

Una vez encontrados los diseños con los parámetros estimados, se eligen distribuciones a priori para estos, debido a que las estimaciones de los parámetros fueron positivas con cualquier función de covarianza tratada, se consideran distribuciones a priori donde éstas toman valores en un subconjunto de los reales positivos o distribuciones donde su dominio sea un conjunto positivo. En este artículo se utilizan tres distribuciones a priori distintas.

La primera distribución a priori pensada para cada uno de los parámetros fue la distribución Uniforme, debido a la facilidad que tiene esta distribución para la elicitación de los extremos del intervalo de variación de dicha distribución, es decir se puede averiguar entre qué valor mínimo y máximo se encuentra cada parámetro. Para propósitos ilustrativos, se tomaron como extremos del intervalo de la distribución uniforme, los extremos de los intervalos de confianza bootstrap, a un 95 % de confianza, dados en la Tabla 4.

**Tabla 4.** Intervalos de confianza bootstrap

	Tipo de función de covarianza		
	Exponencial	Kernel 1	Cauchy 1
$\hat{k}_a$	(5.095, 5.245)	(5.106, 5.237)	(5.109, 5.251)
$\hat{k}_e$	(0.1454, 0.1481)	(0.1451, 0.1482)	(0.1453, 0.1482)
$\hat{\sigma}^2$	(0.010, 0.0413)	(0.014, 0.043)	(0.642, 0.842)
$\hat{\rho}$	(0.437, 0.637)	(0.642, 0.842)	(0.703, 0.903)

Luego de obtener los intervalos de confianza bootstrap, para cada parámetro y cada función de covarianza, se construye la distribución a priori uniforme como ya antes se ha mencionado, este procedimiento que se utilizó sirve para justificar de manera clara y precisa la escogencia de los extremos de la distribución uniforme, en la siguiente tabla se muestra la a priori uniforme para los parámetros.

**Tabla 5.** Distribución Uniforme para cada uno de los parámetros de acuerdo al tipo de función de covarianza

	Tipo de función de covarianza		
	Exponencial	Kernel 1	Cauchy 1
$\hat{k}_a$	U(5.095, 5.245)	U(5.106, 5.237)	U(5.109, 5.251)
$\hat{k}_e$	U(0.1454, 0.1481)	U(0.1451, 0.1482)	U(0.1453, 0.1482)
$\hat{\sigma}^2$	U(0.010, 0.0413)	U(0.014, 0.043)	U(0.642, 0.842)
$\hat{\rho}$	U(0.437, 0.637)	U(0.642, 0.842)	U(0.703, 0.903)

Teniendo en cuenta las distribuciones a priori  $\pi_1(k_a)$ ,  $\pi_2(k_e)$ ,  $\pi_3(\sigma^2)$ ,  $\pi_4(\rho)$ , que son las distribuciones uniforme expuestas en la tabla 5, se considera la distribución a priori conjunta uniforme como el producto de estas cuatro distribuciones a priori, es decir:

$$\pi(\Theta) = \pi_1(k_a) \cdot \pi_2(k_e) \cdot \pi_3(\sigma^2) \cdot \pi_4(\rho) \quad (29)$$

donde  $\Theta^T = (k_a, k_e, \sigma^2, \rho)$  y además se observó que  $\sigma^2$  no tiene ninguna influencia en el criterio (ver colorario 2.1 y la ecuación (16)), esto quiere decir que si se supone una a priori para  $\sigma^2$  el valor de la integral sobre este parámetro es igual a 1, por ende, la distribución conjunta en (29) queda especificada de la siguiente forma:

$$\pi(\Theta) = \pi_1(k_a) \cdot \pi_2(k_e) \cdot \pi_3(\rho) \quad (30)$$

Se maximiza la función dada en (18) usando un programa elaborado por los autores en el software estadístico R [23]. Allí se tiene en cuenta cada una

de las diferentes distribuciones a priori para hallar, vía el método de Monte Carlo y optimización, los diseños óptimos y las eficiencias reportadas entre paréntesis en las tablas 8, 9 y 10.

Adicionalmente, con el fin de determinar el efecto que tiene la elección de la distribución a priori, se consideró la distribución Gamma  $(\alpha, \beta)$  y la distribución Lognormal  $(\mu, s)$ , como distribuciones factibles para cada uno de los elementos del vector de parámetros  $\Theta$ . Para los parámetros de estas distribuciones se determinaron usando el procedimiento descrito a continuación:

Se supone que la media de la distribución uniforme de la tabla 5 es igual a la media de la distribución Gamma y la varianza de la uniforme igual a la varianza de la Gamma (ver Tabla 6), lo anterior conduce al siguiente sistema de ecuaciones:

**Tabla 6.** Sistemas de ecuaciones cuando se iguala la media de la distribución uniforme a la media de la distribución Gamma y la varianza de la Uniforme a la varianza de la Gamma

Tipo de función de covarianza	Parámetros		
	$k_a$	$k_e$	$\rho$
Exponencial	$\begin{cases} 5.170000 = \alpha\beta \\ 0.001875 = \alpha\beta^2 \end{cases}$	$\begin{cases} 0.14680000 = \alpha\beta \\ 0.00000059 = \alpha\beta^2 \end{cases}$	$\begin{cases} 0.537000 = \alpha\beta \\ 0.003333 = \alpha\beta^2 \end{cases}$
Kernel 1	$\begin{cases} 5.1715000 = \alpha\beta \\ 0.0014300 = \alpha\beta^2 \end{cases}$	$\begin{cases} 0.14669000 = \alpha\beta \\ 0.00000084 = \alpha\beta^2 \end{cases}$	$\begin{cases} 0.742000 = \alpha\beta \\ 0.003333 = \alpha\beta^2 \end{cases}$
Cauchy 1	$\begin{cases} 5.1800000 = \alpha\beta \\ 0.0016803 = \alpha\beta^2 \end{cases}$	$\begin{cases} 0.14679000 = \alpha\beta \\ 0.00000069 = \alpha\beta^2 \end{cases}$	$\begin{cases} 0.803000 = \alpha\beta \\ 0.003333 = \alpha\beta^2 \end{cases}$

**Tabla 7.** Sistemas de ecuaciones cuando se iguala la media de la distribución uniforme a la media de la distribución Lognormal y la varianza de la Uniforme a la varianza de la Lognormal

Tipo de función de covarianza	Parámetros		
	$k_a$	$k_e$	$\rho$
Exponencial	$\begin{cases} 5.170000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.001875 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$	$\begin{cases} 0.14680000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.00000059 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$	$\begin{cases} 0.537000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.003333 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$
Kernel 1	$\begin{cases} 5.1715000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.0014300 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$	$\begin{cases} 0.14669000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.00000084 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$	$\begin{cases} 0.742000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.003333 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$
Cauchy 1	$\begin{cases} 5.1800000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.0016803 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$	$\begin{cases} 0.14679000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.00000069 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$	$\begin{cases} 0.803000 = e^{(\mu+\frac{\sigma^2}{2})} \\ 0.003333 = (e^{\sigma^2} - 1) e^{2\mu+\sigma^2} \end{cases}$

Resolviendo las ecuaciones de la Tabla 6, se obtienen las siguientes densidades:

Para la función de covarianza Exponencial

$$\begin{aligned} k_a &\sim \text{Gamma}(14255.41, 0.0003626692) \\ k_e &\sim \text{Gamma}(36005.36, 0.000004077171) \\ \rho &\sim \text{Gamma}(86.5107, 0.006207325) \end{aligned}$$

Para la función de covarianza Kernel 1

$$\begin{aligned} k_a &\sim \text{Gamma}(18701.3, 0.0002765316) \\ k_e &\sim \text{Gamma}(25535.4, 0.0009179732) \\ \rho &\sim \text{Gamma}(165.1692, 0.004492363) \end{aligned}$$

Para la función de covarianza Cauchy 1

$$\begin{aligned} k_a &\sim \text{Gamma}(15968.5, 0.0003243887) \\ k_e &\sim \text{Gamma}(31171.65, 0.000004709088) \\ \rho &\sim \text{Gamma}(193.4427, 0.0041511) \end{aligned}$$

Una vez obtenidas las distribuciones a priori  $\pi_1(k_a)$ ,  $\pi_2(k_e)$ ,  $\pi_3(\rho)$  para cada parámetro y cada función de covarianza, se considera la distribución a priori conjunta Gamma como el producto de estas tres distribuciones a priori, ver la expresión (30)

Con esta nueva distribución a priori conjunta se encuentran los diseños D-óptimos y las eficiencias reportadas entre paréntesis en las tablas 8, 9 y 10 usando el mismo procedimiento detallado con la distribución Uniforme en la página 10.

Posteriormente, también se considera que la media de la distribución uniforme sea igual a la

media de la distribución Lognormal y la varianza de la uniforme igual a la varianza de la distribución Lognormal, como se muestran a continuación:

Solucionando las ecuaciones de la Tabla 7, se encuentra que:

Para la función de covarianza Exponencial

$$\begin{aligned} k_a &\sim \text{Lognormal}(1.64284, 0.00007014604) \\ k_e &\sim \text{Lognormal}(-1.91870, 0.00002777331) \\ \rho &\sim \text{Lognormal}(-0.62750, 0.0114929) \end{aligned}$$

Para la función de covarianza Kernel 1

$$\begin{aligned} ka &\sim \text{Lognormal}(1.64314, 0.00005347057) \\ ke &\sim \text{Lognormal}(-1.91945, 0.00003916083) \\ \rho &\sim \text{Lognormal}(-0.3014241, 0.006036144) \end{aligned}$$

Para la función de covarianza Cauchy 1

$$\begin{aligned} k_a &\sim \text{Lognormal}(1.64477, 0.00006262134) \\ k_e &\sim \text{Lognormal}(-1.91877, 0.00003207979) \\ \rho &\sim \text{Lognormal}(-0.2219787, 0.005156173) \end{aligned}$$

Una vez obtenidas las distribuciones a priori  $\pi_1(k_a)$ ,  $\pi_2(k_e)$ ,  $\pi_3(\rho)$  para cada parámetro, se considera la distribución a priori conjunta Lognormal para cada función de covarianza como el producto de estas tres distribuciones a priori, ver la expresión (30)

Finalmente, se hallan los diseños D-óptimos y las eficiencias reportadas entre paréntesis en las tablas 8, 9 y 10 usando la nueva distribución a priori.

**Tabla 8.** Diseños encontrados utilizando la función de covarianza exponencial y se reporta el valor de la eficiencia del respectivo diseño con relación al diseño óptimo obtenido con la distribución uniforme, en paréntesis.

Puntos de soporte	Función covarianza Exponencial		
	Uniforme	Gamma	Lognormal
2	[ 0.198 2.674 ]	[ 0.198 2.675 ] ( 1 )	[ 0.198 2.674 ] ( 1 )
3	[ 0.000 0.097 6.523 ]	[ 0.000 0.097 6.523 ] ( 1 )	[ 0.000 0.097 6.522 ] ( 1 )
4	[ 0.000 0.098 4.915 7.934 ]	[ 0.000 0.098 4.912 7.933 ] ( 1 )	[ 0.000 0.098 4.912 7.932 ] ( 1 )

**Tabla 9.** Diseños encontrados utilizando la función de covarianza kernel 1 y se reporta el valor de la eficiencia del respectivo diseño con relación al diseño óptimo obtenido con la distribución uniforme, en paréntesis.

Puntos de soporte	Función covarianza Kernel 1		
	Distribución a priori conjunta		
	Uniforme	Gamma	Lognormal
2	[ 0.190 5.030 ]	[ 0.178 0.753 ] ( 0.9102823 )	[ 0.190 5.029 ] ( 1 )
3	[ 0.000 0.103 6.570 ]	[ 0.178 0.763 3.908 ] ( 0.9364363 )	[ 0.000 0.103 6.569 ] ( 1 )
4	[ 0.000 0.104 5.246 8.849 ]	[ 0.000 0.098 0.787 5.144 ] ( 0.9848113 )	[ 0.000 0.104 5.245 8.847 ] ( 1 )

**Tabla 10.** Diseños encontrados utilizando la función de covarianza Cauchy 1 y se reporta el valor de la eficiencia del respectivo diseño con relación al diseño óptimo obtenido con la distribución uniforme, en paréntesis.

Puntos de soporte	Función covarianza Cauchy 1		
	Distribución a priori conjunta		
	Uniforme	Gamma	Lognormal
2	[ 0.193 3.606 ]	[ 0.188 0.648 ] ( 0.9380847 )	[ 0.193 3.605 ] ( 1 )
3	[ 0.000 0.093 6.538 ]	[ 0.000 0.125 0.721 ] ( 0.9028867 )	[ 0.000 0.093 6.537 ] ( 1 )
4	[ 0.000 0.094 4.969 8.707 ]	[ 0.000 0.054 0.134 0.747 ] ( 0.9071434 )	[ 0.000 0.094 4.968 8.705 ] ( 1 )

Para hacer las comparaciones se toma como referencia los diseños hallados por medio de la distribución a priori Uniforme, de esta forma, se observa que si se considera la función de covarianza exponencial de la Tabla 8, las eficiencias de los diseños obtenidos con las diferentes distribuciones a priori son iguales a 1, lo que quiere decir que a la hora de estimar el vector de parámetros se obtienen los mismos resultados si se utiliza cualesquiera de los tres diseños hallados con sus respectivas a priori. Ahora si se considera la función de covarianza Kernel 1 de la Tabla 9, se observa que las eficiencias de los diseños hallados cuando se usa la distribución a priori Gamma dieron los siguientes resultados 0.9102, 0.9364 y 0.9848 para diseños de dos, tres y cuatro puntos de soporte respectivamente, mientras que cuando se emplea la distribución a priori Lognormal las eficiencias siguen siendo iguales a 1. Por último, si se trabaja con la función de covarianza Cauchy 1 de la Tabla 10, se nota que las eficiencias “cuando se emplea la distribución a priori Gamma” fueron de 0.938,

0.9028 y 0.9007 para diseños de dos, tres y cuatro puntos de soporte y las eficiencias para cuando se usa la distribución Lognormal también dieron iguales a 1.

## CONCLUSIONES

De acuerdo al procedimiento propuesto para encontrar diseños D-óptimos bayesianos se evidencia que si hay efecto a la hora de considerar distribuciones a priori en la obtención de los diseños, aunque este efecto depende de la función de covarianza con la que se esté trabajando, puesto que si se considera la función de covarianza exponencial no hay efecto en considerar distribuciones a priori para obtener los diseños óptimos. Pero si se tiene en cuenta las funciones de covarianza Kernel 1 o Cauchy 1 se observa diferencia entre los diseños hallados mediante la distribución a priori Uniforme y la distribución a priori Gamma, esto también se ve reflejado en los valores de las eficiencias.

## REFERENCIAS

- [1] J. Kiefer, “Optimum experimental designs”, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 272-319, 1959.
- [2] J. Kiefer and J. Wolfowitz, “The equivalence of two extremum problems”, *Canadian Journal of Mathematics*, vol. 12, no. 5, pp. 363-365, 1960.
- [3] V. Fedorov, “Theory of Optimal Experiments”, *Academic Press*, 1972.
- [4] F. Pukelsheim, “Optimal design of experiments”, *SIAM*, vol. 50, 1993.
- [5] G. E. Box and H. Lucas, “Design of experiments in non-linear situations”, *Biometrika*, vol. 46, no. 1/2, pp. 77-90, 1959.
- [6] G. E. Box and W. G. Hunter, “Sequential design of experiments for nonlinear models. In Proceedings IBM Scientific Computing Symposium: Statistics”, pp. 113-137, 1965.
- [7] H. Chernoff, “Locally optimal designs for estimating parameters”, *Annals of Mathematical Statistics*, no. 24, pp. 586-602, 1953.
- [8] I. Ford, B. Torsney, and C. F. J. Wu, “The use of a canonical form in the construction of locally optimal designs for nonlinear problems”, *Annals of Mathematical Statistics*, vol. 54, no. 2, pp. 569-583, 1992.
- [9] T. E. O’Brien, “A note on quadratic designs for nonlinear regression models”, *Biometrika*, vol. 79, no. 4, pp. 847-849, 1992.
- [10] H. Dette, V. B. Melas, and W. K. Wong, “Optimal design for goodness-of-fit of the Michaelis Menten enzyme kinetic function”, *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1370-1381, 2005.
- [11] L. Pronzato and É. Walter, “Robust experiment design via stochastic approximation”, *Mathematical Biosciences*, vol. 75, no. 1, pp. 103-120, 1985.
- [12] K. Chaloner and K. Larntz, “Optimal bayesian design applied to logistic regression experiments”, *Journal of Statistical Planning and Inference*, no. 21, pp. 191-208, 1989.
- [13] A. Pázman, “Criteria for optimal design of small-sample experiments with correlated observations”, *Kybernetika*, vol. 43, no. 4, pp. 453-462, 2007.
- [14] D. Ucinski and A. Atkinson, “Experimental design for time-dependent models with correlated observations”, *Studies in Nonlinear Dynamics and Econometrics*, vol. 8, no. 2, 2004.
- [15] M. Amo-Salas, J. López-Fidalgo, and V. López-Ríos, “Optimal designs for two nested pharma-cokinetic models with correlated observations”, *Communications in Statistics-Simulation and Computation*, vol. 41, no. 7, pp. 944-963, 2012.
- [16] S. Campos-Barreiro and J. López-Fidalgo, “Doptimal experimental designs for a growth model applied to a holstein-friesian dairy farm”, *Statistical Methods and Applications*, vol. 24, no. 3, pp. 491- 505, 2015.
- [17] D. A. Ratkowsky, “Nonlinear Regression Modeling: A Unified Practical Approach”, *Marcel Dekker*, 1983.
- [18] A. Atkinson, A. Donev, and R. Tobias, “Optimum experimental designs, with SAS”, *Oxford University Press*, vol. 34, 2007.
- [19] W. G. Müller, L. Pronzato, J. Rendas, and H. Waldl, “Efficient prediction designs for random fields”, *Applied stochastic models in business and industry*, vol. 31, no. 2, pp. 178-194, 2015.

[20] A. Zhigljavsky, H. Dette, and A. Pepelyshev, “A new approach to optimal design for linear models with correlated observations”, *Journal of the American Statistical Association*, vol. 105, no. 491, pp. 1093-1103, 2010.

[21] S. Huet, A. Bouvier, M.-A. Poursat, and E. Jolivet, “Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples”, *Springer Science and Business Media*, 2004.

[22] C. D. Correa Álvarez, “Búsqueda de diseños cuasi-óptimos eficientes a partir de un diseño D-óptimo para observaciones correlacionadas espacialmente”, *Universidad Nacional de Colombia Sede Medellín*.

[23] R Core Team. R: “A Language and Environment for Statistical Computing”, *R Foundation for Statistical Computing, Vienna, Austria*, 2018.

[24] S. E. Rosenbaum, “Basic pharmacokinetics and pharmacodynamics: An integrated textbook and computer simulations”, *John Wiley and Sons*, 2016.

- Se selecciona una muestra aleatoria con reemplazo de tamaño  $n$  de  $\{\tilde{\varepsilon}_i\}$  denotado por  $\{\tilde{\varepsilon}_i^*\}_i^n$

- Se halla  $Y_i^* = \eta(t_i, \hat{\theta}) + \tilde{\varepsilon}_i^*$

- Se encuentra  $\hat{\Theta}^*$  tal que maximiza a

$$l(\Theta^*) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_\beta|) - \frac{1}{2} \left[ (\mathbf{Y}^* - \eta(\mathbf{t}, \theta))^T \Sigma_\beta^{-1} (\mathbf{Y}^* - \eta(\mathbf{t}, \theta)) \right]$$

- El estimador bootstrap es  $\lambda(\hat{\Theta}^*)$ , es decir, se evalúa  $\lambda(\Theta)$  en  $\hat{\Theta}^*$  donde  $\lambda(\Theta)$  es una función que depende del vector de parámetros  $\Theta$  y es la función que se desea estimar o encontrar un intervalo de confianza.

- Ahora, si el procedimiento anterior se repite  $B$  veces, es decir, se realizan  $B$  simulaciones bootstrap, entonces se tiene que:

$$\lambda^{*i} = \lambda(\hat{\Theta}^{*i}), i = 1, 2, \dots, B$$

que es una muestra de tamaño  $B$  de estimaciones bootstrap de  $\lambda$ .

- Para hallar intervalos de confianza para  $\lambda(\Theta)$  en [21] recomiendan usar  $B \approx 200$ .

- Se encuentran los valores de

$$\hat{T}_i^* = \frac{\lambda^{*i} - \hat{\lambda}}{S_{\hat{\lambda}^*}}$$

donde  $\hat{\lambda} = \lambda(\hat{\Theta})$  y la distribución de  $\hat{\lambda}^*$  que se halla a partir de la distribución empírica de los  $\lambda^{*i} = \lambda(\hat{\Theta}^{*i}), i = 1, 2, \dots, B$  dada por:

$$\hat{\lambda}^* = \sum_{i=1}^B \frac{\lambda^{*i}}{B}$$

- A partir de los valores de los cuantiles de orden  $\alpha/2 : b_{\alpha/2}$  y de orden  $1 - \alpha/2 : b_{1-\alpha/2}$  de  $\hat{T}^* = \{\hat{T}_i^*\}_1^B$  se puede hallar un intervalo de confianza bootstrap para  $\lambda = \lambda(\Theta)$ , del  $(1 - \alpha) \times 100\%$  dado por:

## ANEXO

### Intervalos de confianza bootstrap

A continuación, se da una serie de pasos para construir intervalos de confianza Bootstrap enunciado en [21] para modelos de regresión no lineal para cuando  $\varepsilon \sim N_N(0, \sigma^2 I)$ , se adapta para el caso cuando:  $\varepsilon \sim N_N(0, \Sigma\beta)$ .

- Se calcula  $\hat{\Theta}$  el estimador para  $\Theta$  vía máxima verosimilitud.

- Se halla  $\eta(t_i, \hat{\theta}) = \hat{Y}_i$ , se calcula  $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$  de esta forma  $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \hat{\varepsilon}_\bullet$  donde

$$\hat{\varepsilon}_\bullet = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i$$

$$\left( \hat{\lambda}^* - b_{1-\alpha/2} S_{\hat{\lambda}^*}, \hat{\lambda}^* - b_{\alpha/2} S_{\hat{\lambda}^*} \right)$$

En general  $b_{q/2}$  se halla a partir de los valores bootstrap de  $\hat{T}_i^*$   $\forall i = 1, 2, \dots, B$  ordenados de menor a mayor y  $\hat{T}^{*,(b)}$  denota la posición ordenada en que se encuentra los valores de  $\hat{T}^*$

, de esta forma  $b_\alpha$  es  $\hat{T}^{*,(q\alpha)}$  con  $q_\alpha$  el entero más pequeño tal que  $q_\alpha/B$  es mayor o igual a  $\alpha$ .

Por ejemplo, para  $\alpha = 0.05$  y  $B = 199$  se halla  $b_{\alpha/2} = b_{0.025} = \hat{T}^{*,(5)}$  y  $b_{1-\alpha/2} = b_{0.975} = \hat{T}^{*,(195)}$ .