





Diseños D_π -óptimos para modelos no lineales heteroscedásticos: Un estudio de robustez

 Catalina Patiño-Bustamante¹ y  Víctor López-Ríos²

Recepción: 17-02-2019 | Aceptación: 06-05-2020 | En línea: 19-06-2020

MSC: 62K05

doi: 10.17230/ingciencia.16.31.4

Resumen

Los diseños óptimos son utilizados para determinar las mejores condiciones donde se debe realizar un experimento para obtener ciertas propiedades estadísticas. En los modelos no lineales heteroscedásticos donde la varianza es una función de la media, el criterio de optimalidad depende de la elección de un valor local para los parámetros del modelo. Una forma de evitar esta dependencia es considerar una distribución a priori para el vector de parámetros del modelo e incorporarla en el criterio de optimalidad que se va a optimizar. En este artículo se consideran diseños D -óptimos en modelos no lineales heteroscedásticos cuando se incorpora una distribución a priori asociada a los parámetros del modelo. Se extiende el teorema de equivalencia al considerar el efecto de la distribución a priori. Se propone una metodología para la construcción de distribuciones a priori discretas y continuas. Se muestra, con un ejemplo, cómo a partir de las distribuciones construidas se pueden encontrar diseños óptimos con mayor número de puntos experimentales que los obtenidos con un valor local. La eficiencia

¹ Universidad Nacional de Colombia, catpatinobus@unal.edu.co, Medellín, Colombia.

² Universidad Nacional de Colombia, vilopez@unal.edu.co, Medellín, Colombia.

de los diseños hallados es muy competitiva comparada con los diseños óptimos locales. Adicionalmente se consideran distribuciones a priori de una familia de escala, y se muestra que los diseños hallados son robustos a la elección de la distribución a priori elegida de esta familia.

Palabras clave: Diseños óptimos; matriz de información; teorema de equivalencia; distribuciones a priori; modelos heteroscedásticos.

D_π -optimal Designs for Heteroscedastic Nonlinear Models: A Robustness Study

Abstract

Optimal designs are used to determine the best conditions where an experiment should be performed to obtain certain statistical properties. In heteroscedastic nonlinear models where variance is a function of the mean, the optimality criterion depends on the choice of a local value for the model parameters. One way to avoid this dependency is to consider an a priori distribution for the vector of model parameters and incorporate it into the optimality criterion to be optimized. This paper considers D -optimal designs in heteroscedastic nonlinear models when a prior distribution associated with the model parameters is incorporated. The equivalence theorem is extended by considering the effect of the prior distribution. A methodology for the construction of discrete and continuous prior distributions is proposed. It is shown, with an example, how optimal designs can be found from the constructed distributions with a greater number of experimental points than those obtained with a local value. The efficiency of the designs found is very competitive compared to the optimal local designs. Additionally, prior distributions of a scale family are considered, and it is shown that the designs found are robust to the choice of the prior distribution chosen from this family.

Keywords: Optimal designs; information matrix; equivalence theorem; prior distribution; heteroscedastic models.

1 Introducción

Los diseños óptimos permiten utilizar de manera eficiente los recursos requeridos para realizar un experimento en las áreas de investigación donde se desea obtener alguna información de interés mediante el uso de la estadística. De este modo, buscan las mejores condiciones donde se debe realizar un experimento con el fin de que se alcancen algunas propiedades estadísticas, por ejemplo: minimizar la varianza de los estimadores de los

parámetros, minimizar el volumen del elipsoide de confianza asociado a los parámetros del modelo, entre otros. La búsqueda de estos diseños requiere del cumplimiento de los siguientes supuestos: independencia, normalidad y homogeneidad de la varianza del término de error del modelo. En la literatura existen trabajos donde se proponen métodos para hallar diseños óptimos cuando los supuestos anteriores no se cumplen. Por ejemplo, en el caso donde la varianza del error no es constante hay varios aportes para la búsqueda de diseños D -óptimos tanto para modelos lineales como no lineales (Dette y Müller, 2013 [1], Dette y Wong, 1999 [2], Downing, Fedorov y Leonov, 2001 [3], Gaviria y López-Ríos, 2014 [4], Burridge y Sebastiani, 1994 [5], Atkinson y Cook, 1995 [6], Dette y Wong, 1996 [7], Fang y Wiens, 2000 [8]). En todos los trabajos consultados el criterio de optimalidad propuesto depende de la elección de un valor local para los parámetros del modelo. Una forma de evitar esta dependencia es considerar una distribución a priori para el vector de parámetros del modelo e incorporarla en el criterio de optimalidad que se va a optimizar. En el caso de modelos lineales, existen trabajos en esta dirección (Duarte y Wong, 2015 [9], Abebe, Tan, Van Breukelen, Serroyen y Berger, 2014 [10], Burghaus and Dette [11], Buzoianu y Kadane (2019) [12], Braess y Dette [13], Atkinson y Cook, 1995 [6], Woods, Overstall, Adamou y Waite, 2017 [14], Goudarzi, 2019 [15], Bodunwa y Fasoranbaku, 2020 [16]). Mosquera y López-Ríos, 2019 [17] investigan el efecto de diferentes distribuciones a priori en los diseños D -óptimos Bayesianos en el caso de un modelo no lineal donde el supuesto de incorrelación no se cumpla. Adicionalmente, se han publicado recientemente trabajos relacionados con programación para la obtención de diseños D -óptimos Bayesianos (ver Duarte y Wong, 2015 [18] y Abebe et. al. 2017[19]).

En este artículo se investiga el comportamiento de los diseños D -óptimos en modelos no lineales heteroscedásticos cuando se incorpora una distribución a priori asociada a los parámetros del modelo con el fin de dar respuesta a las siguientes preguntas de investigación: ¿cómo se afecta el número de puntos de soporte del diseño D -óptimo con respecto a la elección de la distribución a priori? y ¿qué características deben tener las distribuciones a priori con el fin de que los diseños D -óptimos obtenidos sean robustos a la elección de esta distribución?.

En la sección 2 se dan algunos elementos básicos sobre los diseños óptimos considerando el modelo no lineal heteroscedástico donde se modela la varianza como una potencia de la media. Se presenta la matriz de información para este modelo, se define el criterio de optimalidad y se enuncia el teorema general de equivalencia. Luego, se define el criterio D_π -optimalidad y se enuncia el teorema de equivalencia para este criterio. Finalmente, se define la D_π -eficiencia, la cual permite comparar los diferentes diseños obtenidos. En la sección 3 se realiza la descripción de la metodología propuesta para el modelo considerado. En la sección 4 se halla el diseño D -óptimo local para el ejemplo mencionado. Luego, para evitar la dependencia de este vector, se construyen distribuciones a priori discretas y continuas, lo cual se explica en las subsección 4.4 en las secciones 4.4.1 y 4.4.2 se obtienen diseños D_π -óptimos para el modelo inicial con tres y cuatro puntos de soporte y se comparan con el diseño D -óptimo local mediante el cálculo de la eficiencia. Finalmente, en la sección 4.4.3, para estudiar la dependencia del diseño a la elección de la a priori, se realiza un estudio vía simulación, donde se toman diferentes distribuciones que pertenecen a la familia de escala de modo que también pertenezca la distribución a priori continua construída a partir del vector de parámetros local y se puede concluir que el diseño es robusto a la elección de la distribución a priori.

2 Conceptos preliminares

Se considera el siguiente modelo donde la variable respuesta Y está relacionada con las variables explicativas $\mathbf{x} = (x_1, \dots, x_k)^T$ a través de la siguiente expresión:

$$Y_i = Y(\mathbf{x}_i) = \eta(\mathbf{x}_i, \alpha) + \epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

donde se dispone de n corridas experimentales, \mathbf{x} varía en un espacio compacto $\chi \subseteq \mathbb{R}^k$, dotado de una σ -álgebra \mathcal{B} , $\alpha \in \mathbb{R}^m$, $\eta(\mathbf{x}_i, \alpha)$ es una función no lineal en α y los ϵ_i son independientes, tienen distribución normal con media cero y varianza $Var(\epsilon_i) = S(\mathbf{x}_i, \alpha, \beta) = \sigma^2 (\eta(\mathbf{x}_i, \alpha))^{2\tau}$, donde $\beta = (\sigma^2, \tau)$, $\sigma \in \mathbb{R}^+$ y $\tau \in \mathbb{R}$ (Ritz, 2008 [20]).

Asociado al modelo (1), se define un diseño ξ como:

$$\xi = \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_d \\ w_1 & \dots & w_d \end{bmatrix},$$

donde $w_j = \xi(\mathbf{x}_j)$, $j = 1, \dots, d$ y ξ es una medida de probabilidad definida en \mathcal{B} , tal que ξ tiene soporte finito, $\text{Sop}(\xi) = \{\mathbf{x}_j \in \mathcal{X} \mid \xi(\mathbf{x}_j) > 0\} = \{\mathbf{x}_1, \dots, \mathbf{x}_d\}$, d es el número de puntos de soporte de ξ y las observaciones $Y(\mathbf{x})$ se hacen en $\mathbf{x}_1, \dots, \mathbf{x}_d$ con pesos asociados aproximadamente proporcionales a w_1, \dots, w_d . Si $0 < w_j \leq 1$, se dice que el diseño es continuo o aproximado (López-Ríos, Ramos, 2007) [21].

2.1 Matriz de información

Al considerar la estructura de la varianza de los errores del modelo definida por $\text{Var}(\epsilon_i) = \sigma^2 (\eta(\mathbf{x}_i, \alpha))^{2\tau}$, interesa hallar la expresión para la matriz de información de Fisher asociada al diseño ξ , de modo que ésta incorpore dicha estructura.

Ahora, para el modelo (1) donde $\text{Var}(\epsilon) = S(\mathbf{x}, \alpha, \beta)$ es una función positiva, se tiene que la matriz de información de Fisher para el diseño aproximado es (Downing, Fedorov y Leonov, 2001 [3])

$$M(\xi, \theta) = \int_{\mathcal{X}} I(\mathbf{x}, \theta) \xi(d\mathbf{x}),$$

donde

$$I(\mathbf{x}, \theta) = \frac{1}{S(\mathbf{x}, \theta)} \frac{\partial \eta(\mathbf{x}, \alpha)}{\partial \theta} \frac{\partial \eta(\mathbf{x}, \alpha)}{\partial \theta^T} + \frac{1}{2S(\mathbf{x}, \theta)^2} \frac{\partial S(\mathbf{x}, \theta)}{\partial \theta} \frac{\partial S(\mathbf{x}, \theta)}{\partial \theta^T}, \quad (2)$$

para $\theta = (\alpha, \beta)^T$.

De acuerdo a la expresión anterior se observa que la matriz de información para este modelo se puede escribir de la siguiente manera:

$$M(\xi, \theta) = \int_{\mathcal{X}} I(\mathbf{x}, \theta) \xi(d\mathbf{x}) = \sum_{j=1}^d I(\mathbf{x}_j, \theta) w_j = UWU^T + VWV^T, \quad (3)$$

donde

$$U_{(m+2) \times d} = \begin{bmatrix} \uparrow & & \uparrow \\ u_1 & \cdots & u_d \\ \downarrow & & \downarrow \end{bmatrix} \quad V_{(m+2) \times d} = \begin{bmatrix} \uparrow & & \uparrow \\ v_1 & \cdots & v_d \\ \downarrow & & \downarrow \end{bmatrix}$$

$$W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & w_d \end{bmatrix}$$

donde

$$u_j = \frac{1}{\sqrt{S(\mathbf{x}_j, \theta)}} \frac{\partial \eta(\mathbf{x}_j, \alpha)}{\partial \theta} \quad \text{y} \quad v_j = \frac{1}{S(\mathbf{x}_j, \theta)} \frac{\partial S(\mathbf{x}_j, \theta)}{\partial \theta},$$

para $j = 1, \dots, d$.

Gaviria y López-Ríos (2014) [4] muestran, como una consecuencia del resultado anterior, la forma de la matriz de información para el caso en el cual la estructura de la varianza del error es $S(\mathbf{x}, \beta) = \sigma^2 (\eta(\mathbf{x}, \alpha))^{2\tau}$:

$$M(\xi, \theta) = UWU^T + VWV^T,$$

donde

$$u_j = \left[\frac{1}{\sigma \eta(\mathbf{x}_j, \alpha)} \frac{\partial \eta(\mathbf{x}_j, \alpha)}{\partial \theta}, \quad 0, \quad 0 \right]_{(m+2) \times 1}^T$$

$$v_j = \left[\frac{\sqrt{2}\tau}{\eta(\mathbf{x}_j, \alpha)} \frac{\partial \eta(\mathbf{x}_j, \alpha)}{\partial \theta}, \quad \sqrt{2} \ln \eta(\mathbf{x}_j, \alpha), \quad \frac{1}{\sqrt{2}\sigma^2} \right]_{(m+2) \times 1}^T$$

para $j = 1, \dots, d$.

Una vez encontrada la matriz de información de Fisher, interesa hallar el diseño que optimice algún funcional de valor real de esta matriz, denominado criterio de optimalidad.

2.2 Criterios de optimalidad

Como se puede observar en el modelo no lineal considerado en este artículo, donde los errores tienen una distribución normal y su varianza es una función de la media (ver ecuación (1)), la matriz de información depende del vector de parámetros local y por lo tanto el diseño depende de los valores que tomen las componentes de este vector. Una manera de evitar esta dependencia es considerar una distribución a priori para el vector de parámetros. De acuerdo a lo anterior se plantea el criterio D_π -optimalidad de la siguiente manera:

Criterio D_π -optimalidad

$$\phi_{D_\pi}(M(\xi)) = \sum_{\theta \in \Theta} \phi_D(M(\xi, \theta)) \cdot \pi(\theta) = \sum_{\theta \in \Theta} \log |M(\xi, \theta)| \cdot \pi(\theta), \quad (4)$$

donde $\pi(\theta)$ es una distribución de probabilidad discreta. De manera análoga se plantea este criterio para el caso en el cual la distribución π es continua:

$$\phi_{D_\pi}(M(\xi)) = \int_{\theta \in \Theta} \phi_D(M(\xi, \theta)) \cdot \pi(\theta) d\theta = \int_{\theta \in \Theta} \log |M(\xi, \theta)| \cdot \pi(\theta) d\theta. \quad (5)$$

Se dice que un diseño ξ^* es D_π -óptimo si maximiza el criterio $\phi_{D_\pi}(M(\xi))$ dado en las expresiones (4) y (5).

2.3 Teorema de equivalencia

Uno de los objetivos principales de este artículo es extender el teorema de equivalencia de Kiefer y Wolfowitz [22], cuando se consideran distribuciones a priori para el vector de parámetros local en modelos no lineales heteroscedásticos. A continuación se da la definición de derivada direccional y se enuncian los teoremas de equivalencia para el criterio D_π -optimalidad cuando la distribución a priori es discreta y cuando esta distribución es continua.

Definición 1. *Dados dos diseños $\xi, \eta \in \Xi$, se define la derivada direccional de ξ en la dirección de η por:*

$$\Phi(\xi, \eta) := \Phi(M(\xi), M(\eta)) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \{ \phi((1-\epsilon) \cdot M(\xi) + \epsilon \cdot M(\eta)) - \phi(M(\xi)) \}$$

Ahora, se puede probar que la derivada direccional puede expresarse como se indica en el siguiente lema (para detalles de la prueba ver [23])

Lema 1. *Cuando la distribución a priori π es discreta, se cumple que*

$$\Phi_{D_\pi}(M(\xi^*), M(\xi_{\mathbf{x}})) = \sum_{\theta \in \Theta} \text{Tr}(M(\xi_{\mathbf{x}}, \theta) \cdot M^{-1}(\xi^*, \theta)) \cdot \pi(\theta) - p,$$

donde $p = \dim(\Theta)$. En el caso continuo, se tiene que

$$\Phi_{D_\pi}(M(\xi^*), M(\xi_{\mathbf{x}})) = \int_{\theta \in \Theta} \text{Tr}(M(\xi_{\mathbf{x}}, \theta) \cdot M^{-1}(\xi^*, \theta)) \cdot \pi(\theta) d\theta - p.$$

El siguiente teorema es válido tanto para distribuciones a priori discretas como continuas. Los detalles se pueden consultar en [24].

Teorema 1. *Sean $M(\xi, \theta)$ la matriz de información del diseño ξ definida positiva, $\phi_{D_\pi}(M(\xi))$ el criterio D_π -óptimalidad asociado a la distribución a priori π , χ un conjunto compacto, p el número de parámetros del modelo y $\xi_{\mathbf{x}}$ el diseño que asigna toda la probabilidad al punto \mathbf{x} . Si la derivada direccional Φ_{D_π} en ξ^* en la dirección de $\xi_{\mathbf{x}}$ satisface:*

$$\Phi_{D_\pi}(M(\xi^*), M(\xi_{\mathbf{x}})) \leq 0, \quad \forall \mathbf{x} \in \chi,$$

y además se cumple que

$$\Phi_{D_\pi}(M(\xi^*), M(\xi_{\mathbf{x}})) = 0, \quad \text{para todo punto de soporte de } \xi^*,$$

entonces el diseño ξ^* es D_π -óptimo.

2.4 D-eficiencia

Dados un diseño ξ y un diseño ξ^* D_π -óptimo, es posible compararlos mediante la eficiencia:

$$ef_{D_\pi}(\xi) = \frac{\phi_{D_\pi}(\xi)}{\phi_{D_\pi}(\xi^*)}, \quad \text{donde } 0 < ef_{D_\pi}(\xi) \leq 1.$$

La eficiencia muestra qué tan próximo está el diseño ξ de la información dada por el diseño D_π -óptimo ξ^* . Se dice que el diseño ξ es *comparable con* el diseño D_π -óptimo si su eficiencia es cercana a 1.

3 Metodología

A continuación se describe la metodología propuesta para hallar los diseños óptimos heteroscedásticos para el modelo propuesto en la ecuación (1):

1. Se determinan los valores locales de los parámetros a partir de la estimación de los mismos por máxima verosimilitud. Se encuentran los intervalos de confianza vía bootstrap.
2. Con el valor local hallado en el paso anterior, se halla el diseño óptimo que maximice el determinante de la matriz de información dada en la ecuación (3).
3. Se construyen las diferentes distribuciones a priori descritas a continuación:
 - (a) Distribución uniforme discreta para cada parámetro: Se halla al perturbar por una cantidad δ , tanto a la derecha como a la izquierda, cada componente del vector de valores locales, es decir, para cada parámetro θ_i , $1 \leq i \leq r$, se obtienen dos perturbaciones $\theta_i(1-\delta)$ y $\theta_i(1+\delta)$. Luego, se consideran todas las posibles combinaciones de las perturbaciones obtenidas, generando un conjunto de 2^r vectores para cada δ . A este conjunto se le adiciona el vector de parámetros local

y se le asigna a cada uno, una probabilidad de $1/(2^r + 1)$. De este modo, la distribución a priori a considerar dará la misma probabilidad a cada uno de los puntos del espacio parametral.

- (b) Distribución uniforme continua: Para cada θ_i se considera una distribución a priori continua definida en el respectivo intervalo de confianza hallado en el paso 1. La distribución uniforme continua conjunta, π , es el producto de las respectivas distribuciones a priori individuales para cada componente θ_i .
 - (c) Distribución a priori de la familia de escala: Se construye la familia de escala para cada uno de los θ_i , de tal forma que cada distribución uniforme continua considerada en (b) esté incluida en esta familia.
4. Para cada una de las distribuciones a priori dadas en el paso 3, se halla el respectivo diseño D_π -óptimo y se calcula la respectiva eficiencia. Lo anterior se hace implementando un programa en el language estadístico R, [25].

4 Estudio de un caso

En esta sección, se presenta un ejemplo donde se considera un modelo no lineal heteroscedástico y se halla el diseño D_π -óptimo local. Luego, a partir del vector de parámetros local se construyen distribuciones a priori para evitar la dependencia de éste y se comparan los diseños obtenidos con estas distribuciones con el diseño D_π -óptimo local mediante la D_π -eficiencia.

4.1 Descripción de los datos

La metodología anterior se bosqueja a partir de un caso de estudio con datos tomados de Bates & Watts (1988). Los datos corresponden a las concentraciones de residuos de policlorobifenilos (PCB) en una

clase de trucha del lago Cayuga, medidas en 28 peces y se tiene el registro exacto de la edad (en meses) debido a su almacenamiento anual. Todos los peces se cortaron, se molieron, se mezclaron y se tomaron muestras de cinco gramos. Se trataron las muestras y se estimaron los residuos de PCB en partes por millón (ppm) usando cromatografía en columna.

En el gráfico de dispersión, ver Figura 1, se puede observar que a medida que aumenta la edad de los peces también aumenta la variabilidad, por lo que la varianza del error no es constante.

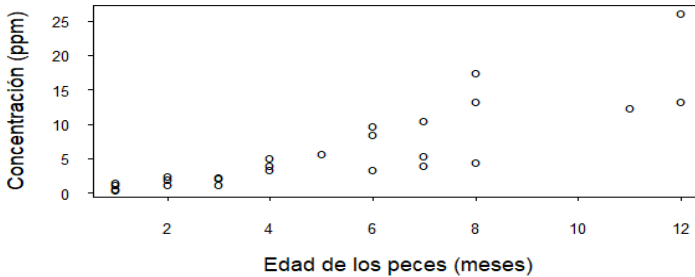


Figura 1: Gráfico de dispersión de la edad de los peces contra la concentración de PCB. Fuente: Elaboración propia.

4.2 Ajuste del modelo

En la Figura (1) se puede observar que no hay una relación lineal entre las variables; por lo cual, se propone ajustar el siguiente modelo no lineal al conjunto de datos:

$$Y_i = \alpha_0 e^{\alpha_1 x_i} + \epsilon_i, \quad \epsilon_i \sim NID(0, \sigma^2) \quad (6)$$

donde α_0 y α_1 son los parámetros a estimar, los errores tienen distribución normal con media cero y varianza constante y son independientes. Ahora, uno de los supuestos que se requieren para hallar los diseños óptimos es que se cumpla la homogeneidad de la varianza de

los errores. Al realizar los gráficos de: residuales contra la edad de los peces, residuales contra los valores ajustados y el test de Levene para la homogeneidad de varianza, se puede concluir que la varianza del error no es constante por lo que se propone considerar la varianza como una función de la media. El modelo propuesto es el siguiente

$$Y_i = \alpha_0 e^{\alpha_1 x_i} + \epsilon_i, \quad \epsilon_i \sim NI(0, S(x_i, \theta)) \tag{7}$$

donde los errores se suponen independientes con distribución normal (NI), media cero y varianza

$$S(x_i, \theta) := S(x_i, \alpha, \beta) = \sigma^2 (\alpha_0 e^{\alpha_1 x_i})^{2\tau}, \quad \text{donde } \alpha = (\alpha_0, \alpha_1), \quad \beta = (\sigma^2, \tau). \tag{8}$$

Ahora, interesa hallar las estimaciones de los parámetros α_0 , α_1 , τ y σ . Para ésto, se utiliza el método de mínimos cuadrados generalizados, mediante la función *gnls* de R [25]. Las estimaciones puntuales obtenidas se muestran en la Tabla (1):

Tabla 1: Estimaciones de los parámetros del modelo usando la función de varianza dada en (7).

$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\tau}$	$\hat{\sigma}^2$
0.9687276	0.2939170	1.115642	0.3733493 ²

Luego, se usará como vector de parámetros local:

$$\hat{\theta}_0 = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\tau}, \hat{\sigma}^2) = (0.9687276, 0.2939170, 1.115642, 0.3733493^2) \tag{9}$$

Con la prueba de razón de verosimilitudes se comprueba que, en efecto, el modelo con función de varianza dada en (8) es mejor que el modelo (6), cuando se supone varianza constante, ver resultados en la Tabla 2:

Tabla 2: Prueba de razón de verosimilitud entre los modelos (6) y (8)

Modelo	gl	AIC	BIC	logVer	Prueba	R. Vero	Valor-P
8	4	114.50	119.53	-53.25			
6	3	141.49	145.26	-67.74	8 vs 6	28.98	< .0001

4.3 Diseño D -óptimo local

Con el valor local θ_0 , se halla el diseño D -óptimo local a partir del uso de dos funciones, la función $DEoptim$ de R, [25], para hallar los valores iniciales y luego, con estos valores se utiliza la función $nlminb$ de R [25] y se obtiene el diseño D -óptimo local siguiente:

$$\xi^* = \begin{bmatrix} 1 & 12 \\ 0.5 & 0.5 \end{bmatrix}.$$

Con el respectivo teorema de equivalencia se verifica que en efecto este diseño es D -óptimo local.

4.4 Diseños D_π -óptimos

En esta sección se obtienen diseños D_π -óptimos para el modelo inicial con tres y cuatro puntos de soporte y se comparan con el diseño D -óptimo local mediante el cálculo de la eficiencia. Se estudia la dependencia del diseño a la elección de la a priori, por medio de un estudio vía simulación, donde se toman diferentes distribuciones que pertenecen a la familia de escala de modo que también pertenezca la distribución a priori continua construída a partir del vector de parámetros local y se puede concluir que el diseño es robusto a la elección de la distribución a priori.

4.4.1 Distribuciones a priori discretas

Para este estudio se van a considerar distribuciones uniformes discretas definidas en los diferentes intervalos generados a partir de las

componentes del vector de parámetros local, los cuales se definen a continuación:

Con el vector de parámetros dado en (9) se generan las distribuciones a priori discretas uniforme de la siguiente manera:

1. Se toman diferentes valores de δ entre 0.40 y 0.95 con incrementos de 0.05, denotado por $\delta = 0.40 : 0.95 : 0.05$, generando 11 valores de δ , con el fin de obtener diferentes perturbaciones de θ_0 .
2. Se multiplica cada componente del vector de parámetros local por $(1 - \delta)$ y $(1 + \delta)$ y se consideran todas las combinaciones posibles de estos valores, generando así un conjunto de 2^4 vectores para cada δ .
3. Al conjunto del numeral anterior se le adiciona el vector de parámetros local, obteniendo un conjunto, Θ_δ , con 17 vectores. Luego, se le asigna una probabilidad $1/17$ a cada vector. De lo anterior, se obtiene una distribución a priori discreta uniforme para cada δ :

$$\pi_\delta(\theta) = \begin{cases} \frac{1}{17}, & \theta \in \Theta_\delta \\ 0, & \text{en otro caso} \end{cases}$$

Con las diferentes distribuciones a priori definidas anteriormente se hallan los diseños D_π -óptimos mostrados en las Tablas (3) y (4). Se puede observar que para perturbaciones de δ de hasta del 0.65 se obtuvieron diseños con tres puntos de soporte, mientras que para perturbaciones del vector de parámetros local mayores al 0.65 se consiguieron diseños con cuatro puntos de soporte.

Tabla 3: Diseños D_π -óptimos para $\delta = 0.40 : 0.65 : 0.05$.

δ	ξ				δ	ξ											
0.40	1	4.991022	12		0.499954	0.000100	0.499946		0.45	1	4.203465	12		0.464016	0.067229	0.468755	
0.50	1	3.957213	12		0.437390	0.111264	0.451347		0.55	1	3.742173	12		0.416861	0.141145	0.441994	
0.60	1	3.556123	12		0.400487	0.162112	0.437401		0.65	1	3.397306	12		0.386931	0.177566	0.435503	

Tabla 4: Diseños D_π -óptimos para $\delta = 0.70 : 0.95 : 0.05$.

δ	ξ				δ	ξ													
0.70	1	2.8580	9.6864	12		0.3661	0.1535	0.0900	0.3904	0.75	1	2.6452	9.9419	12		0.3550	0.1549	0.1156	0.3744
0.80	1	2.4856	10.1450	12		0.3469	0.1579	0.1323	0.3629	0.85	1	2.3599	10.3127	12		0.3405	0.1615	0.1436	0.3543
0.90	1	2.2571	10.4539	12		0.3353	0.1655	0.1517	0.3476	0.95	1	2.1707	10.5744	12		0.3306	0.1695	0.1579	0.3419

Ahora, mediante la D_π -eficiencia, se comparan algunos de los diseños D_π -óptimos obtenidos, que tan competitivos son los diferentes diseños hallados con relación a un diseño de referencia. En la Figura (2) se muestran las D_π -eficiencias considerando como diseño de referencia el diseño óptimo local.

Se observa que la mayoría de las eficiencias son cercanas a uno (ver Figura 2). Luego, los diseños obtenidos con las distribuciones a priori uniformes discretas con las diferentes perturbaciones son casi tan eficientes como el diseño obtenido con el vector de parámetros local.

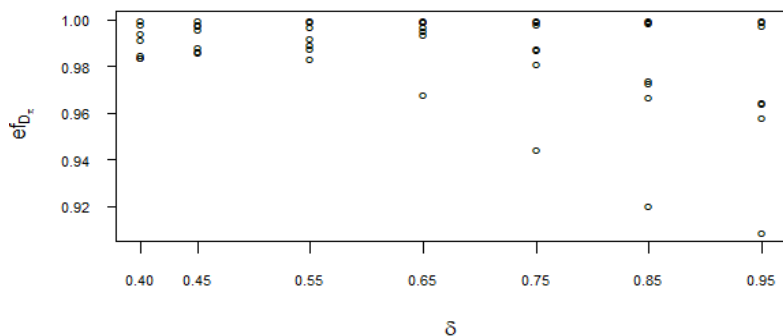


Figura 2: D_π -eficiencia de los diseños mostrados en las Tablas (3) y (4) para diferentes valores de δ

Tabla 5: Comparación de los diseños mediante la D_π -eficiencia para los diseños obtenidos con una perturbación δ tomando como diseño de referencia el diseño D_π -óptimo con perturbación δ_0

$\delta \backslash \delta_0$	0.40	0.45	0.55	0.65	0.75	0.85	0.95
0.40	1	0.998473	0.994301	0.991821	0.984948	0.983906	0.983974
0.45	0.997526	1	0.998129	0.996241	0.988435	0.986692	0.986142
0.55	0.983394	0.997186	1	0.999457	0.992248	0.989513	0.987774
0.65	0.968335	0.993932	0.999545	1	0.999601	0.997368	0.995292
0.75	0.944467	0.981039	0.987230	0.988044	1	0.999435	0.998046
0.85	0.920504	0.967046	0.973236	0.974102	0.998982	1	0.999633
0.95	0.909033	0.958461	0.964231	0.965076	0.997681	0.999601	1
Local	0.999984	0.989810	0.979430	0.975753	0.972353	0.973479	0.975224

La Tabla 5 muestra todas las D_π -eficiencias obtenidas al comparar todos los diseños D_π -óptimos dos a dos y con el diseño D -óptimo local. Se puede observar que al comparar el diseño D -óptimo local con el diseño D_π -óptimo obtenido al tomar una distribución uniforme construida a partir de una perturbación del 95 % se obtuvo una D_π -eficiencia del 97 %, por lo que se puede concluir que el diseño óptimo local es tan eficiente como el diseño D_π -óptimo con $\delta = 0.95$.

4.4.2 Distribuciones a priori continuas En general, cuando se tienen distribuciones a priori continuas, en el proceso de maximización se debe evaluar la expresión dada en (5). Para ello, se utiliza simulación Monte Carlo como se indica a continuación: Se generan NS números aleatorios de la distribución a priori $\pi(\theta)$; por la ley débil de los grandes números, para la distribución $\pi(\theta)$ y para la función $\log |M(\xi, \theta)|$, se tiene que

$$\frac{1}{NS} \sum_{i=1}^{NS} \log |M(\xi, \theta_i)| \xrightarrow{NS \rightarrow \infty} E_{\pi}[\log |M(\xi, \theta)|],$$

donde la notación $a_n \xrightarrow[n \rightarrow \infty]{P} b$, indica que la serie $\{a_n\}$ converge en probabilidad a b y

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

Es decir, $\frac{1}{NS} \sum_{i=1}^{NS} \log |M(\xi, \theta_i)|$ converge en probabilidad al criterio D_{π} -óptimo $\phi_{D_{\pi}}(M(\xi))$. El procedimiento anterior se aplicó para el caso de distribuciones uniforme continuas tomando $NS = 1000$, descritas más adelante.

Mediante simulaciones Bootstrap (Rana, Midi e Imon, [26]), se obtienen los intervalos de confianza para el vector de parámetros local dados en la Tabla (6):

Tabla 6: Intervalos de confianza Bootstrap para las componentes del vector de parámetros local.

	α_0	α_1	τ	σ
2.5 %	$a_1 = 0.48416600$	$a_2 = 0.21304242$	$a_3 = 0.09632032$	$a_4 = 0.09321783$
97.5 %	$b_1 = 1.4118453$	$b_2 = 0.3706811$	$b_3 = 1.7988547$	$b_4 = 2.5091867$

Los números aleatorios de las distribuciones a priori uniformes continuas a partir del vector de parámetros local se generan de la siguiente manera:

1. Se generan números aleatorios de la distribución uniforme continua en $[0, 1]$, vía simulación, mediante la función *runif* de R, [25].
2. Luego, por medio de la transformación $g_i : [0, 1] \longrightarrow [a_i, b_i]$ definida por

$$g_i(x) = a_i + (b_i - a_i)x$$

para $i = 1, 2, 3$ y 4 , se generan números aleatorios de las cuatro distribuciones a priori continuas, $\pi_{[a_1, b_1]}$, $\pi_{[a_2, b_2]}$, $\pi_{[a_3, b_3]}$, $\pi_{[a_4, b_4]}$, para cada una de las cuatro componentes del vector de parámetros local en los intervalos de confianza $[a_1, b_1]$, $[a_2, b_2]$, $[a_3, b_3]$ y $[a_4, b_4]$.

3. Los números aleatorios generados en el numeral 2, serán los números aleatorios de la distribución a priori uniforme continua final definida como la distribución de probabilidad conjunta de las distribuciones definidas en el numeral anterior:

$$\pi(\theta) = \pi(\theta_1, \theta_2, \theta_3, \theta_4) = \pi_{[a_1, b_1]}(\theta_1) \cdot \pi_{[a_2, b_2]}(\theta_2) \cdot \pi_{[a_3, b_3]}(\theta_3) \cdot \pi_{[a_4, b_4]}(\theta_4)$$

donde $\theta \in [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \times [a_4, b_4]$.

Con los números aleatorios generados anteriormente, se evalúa el criterio D_π -optimalidad y se obtiene el diseño D_π -óptimo siguiente:

$$\xi^* = \begin{bmatrix} 1 & 8.36850086 & 12 \\ 0.4930184 & 0.01511487 & 0.4918097 \end{bmatrix}$$

El diseño anterior, sugiere que el 49% de las corridas experimentales se debe realizar en peces de un año, aproximadamente el 49% en peces de doce años y el 1.5%, en peces de ocho años y cuatro meses aproximadamente.

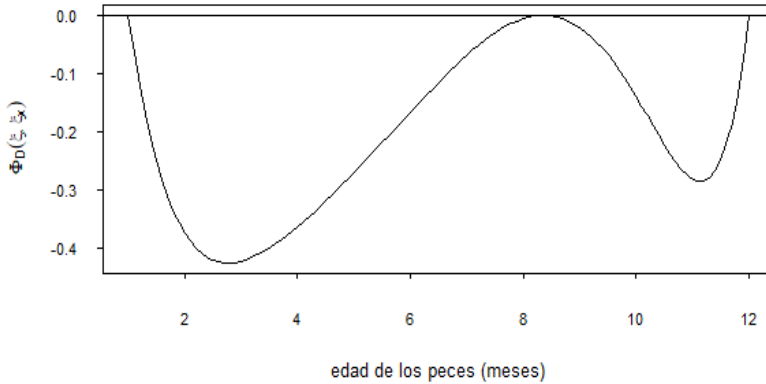


Figura 3: Derivada Direccional del Diseño D_π - óptimo obtenido con 1000 puntos vía simulación.

En la Figura (3) se muestra la derivada direccional para el diseño obtenido con 1000 puntos. De esta gráfica, se puede observar que el diseño obtenido, es D_π -óptimo, ya que la derivada es cero en los puntos de soporte y es estrictamente menor que cero en los demás puntos.

4.4.3 Estudio de robustez a la elección de la distribución a priori

En esta sección se estudia la robustez a la selección de la distribución a priori vía simulación, donde se considera como diseño base el diseño D_π -óptimo obtenido con el vector de parámetros local y una familia de escala donde las distribuciones se construyen a partir de una distribución uniforme continua en $[0, 1]$, donde se fija el extremo izquierdo de los intervalos de confianza de las componentes del vector de parámetros local obtenidos en la sección anterior, a_i , para $i = 1, 2, 3$ y 4 y se toma como parámetro de escala el valor $(b_i - a_i)k$, $k \in \mathbb{R}^+$.

De la sección 4.4.2, el diseño D_π -óptimo obtenido estaba dado por:

$$\xi^* = \begin{bmatrix} 1 & 8.36850086 & 12 \\ 0.4930184 & 0.01511487 & 0.4918097 \end{bmatrix}.$$

Ahora, sean $U_i \sim U_{[0,1]}$, $i = 1, 2, 3$ y 4 , la distribución uniforme continua en $[0, 1]$ obtenida mediante simulaciones en \mathbb{R} , con función de densidad de probabilidad $f_i(x)$, para cada una de las cuatro componentes del vector de parámetros local. Se realiza un procedimiento análogo al descrito en la sección 4.4.2 para construir estas distribuciones.

La familia de escala para el parámetro θ_i considerada en este caso, es la familia que tiene como parámetro de escala $(b_i - a_i)k$, donde k es un número real positivo:

$$\mathcal{F}_i = \left\{ \pi_i : [a_i, b_i] \longrightarrow \mathbb{R} \mid \pi_i(\theta) = \frac{1}{(b_i - a_i)k} \cdot f_i\left(\frac{\theta - a_i}{(b_i - a_i)k}\right) \right\},$$

con f_i la función de densidad de probabilidad de la distribución uniforme $U_{[0,1]}$. Se considera esta familia con el fin de que cada una de las distribuciones uniforme continuas, $\pi_{[a_i, b_i]}$ construidas en el paso 2 de la sección anterior, pertenezca a la familia, \mathcal{F}_i , lo cual se obtiene al tomar $k = 1$ y así poder compararla con otras distribuciones, que también pertenezcan a la familia, \mathcal{F}_i , obtenidas con diferentes valores de k .

Para la construcción de la familia de localización y escala, se siguen los siguientes pasos:

1. En la sección anterior se definieron los intervalos de confianza para α_0, α_1, τ y σ , los cuales se denotaron por: $[a_1, b_1]$, $[a_2, b_2]$, $[a_3, b_3]$ y $[a_4, b_4]$.
2. Luego, se generan 1000 números aleatorios de la distribución uniforme con la función *runif* de \mathbb{R} , [25], para cada una de las cuatro componentes del vector de parámetros local. Denote por U_i los valores aleatorios generados.

3. Sean $Z_i = a_i + k(b_i - a_i)U_i$, las variables aleatorias construidas a partir de las variables U_i . Entonces Z_i tiene f.d.p.

$$\frac{1}{(b_i - a_i)k} \cdot f\left(\frac{x - a_i}{(b_i - a_i)k}\right),$$

con $f(\cdot)$ la función de densidad de la distribución $U_{[0,1]}$.

4. Finalmente, de manera análoga a como se hallaron números aleatorios de las distribuciones a priori continuas, se generan números aleatorios de la distribución a priori uniforme continua final definida como la distribución de probabilidad conjunta de las distribuciones definidas en el numeral anterior:

$$\pi(\theta) = \pi(\theta_1, \theta_2, \theta_3, \theta_4) = \pi_{[a_1, b_1]}(\theta_1) \cdot \pi_{[a_2, b_2]}(\theta_2) \cdot \pi_{[a_3, b_3]}(\theta_3) \cdot \pi_{[a_4, b_4]}(\theta_4)$$

donde $\theta \in [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \times [a_4, b_4]$.

Después de generar los números aleatorios de las distribuciones a priori uniformes continuas, y a partir de un programa elaborado en R, [25], se obtienen los diseños D_π - óptimos, mostrados en la Tabla 7:

Tabla 7: Diseños D_π -óptimos obtenidos para diferentes valores de k .

k	ξ	k	ξ
0.9	$\begin{bmatrix} 1 & 12 \\ 0.5 & 0.5 \end{bmatrix}$	1.1	$\begin{bmatrix} 1 & 5.8507 & 12 \\ 0.4719 & 0.0541 & 0.4740 \end{bmatrix}$
1.2	$\begin{bmatrix} 1 & 4.3692 & 12 \\ 0.4421 & 0.1054 & 0.4525 \end{bmatrix}$	1.3	$\begin{bmatrix} 1 & 3.8663 & 12 \\ 0.4206 & 0.1456 & 0.4338 \end{bmatrix}$

Tabla 8: D_π -eficiencia de los diseños obtenidos de la familia de escala con respecto al diseño D_π -óptimo.

k	$D_\pi - \text{eficiencia}$
0.9	0.9996411
1.1	0.9962667
1.2	0.9866717
1.3	0.9779870

En la Tabla (8) se muestran las D_π -eficiencias que resultan de comparar los diseños obtenidos de la familia de escala, con $k = 0.9, 1.1, 1.2, 1.3$ con respecto al diseño D_π -óptimo. Este último, se halla, con la respectiva a priori que pertenece a esta misma familia tomando $k = 1$. De los resultados arrojados, se puede notar que todas las D_π -eficiencias están por encima del 98 %, permitiendo concluir que los diseños obtenidos con las diferentes distribuciones consideradas pertenecientes a la familia son casi tan eficientes como el diseño obtenido con la distribución de referencia, es decir, cuando $k = 1$. Lo anterior, permite afirmar que el diseño no depende de la elección de la distribución a priori, es decir, el diseño es robusto a la elección de la a priori, al menos si se selecciona la distribución a priori entre las distribuciones investigadas.

5 Conclusiones

En este artículo se extendió el teorema de equivalencia de Kiefer y Wolfowitz [22] al caso heteroscedástico cuando se considera en el criterio de optimalidad información a priori de los parámetros del modelo. Se propuso una metodología para la construcción de distribuciones a priori tanto continuas como discretas, las cuales, a partir de un ejemplo, permitieron incrementar el número de puntos experimentales del diseño de dos hasta cuatro puntos, lo cual es beneficioso para el usuario cuando se necesitan hacer pruebas de bondad de ajuste para el modelo que se está considerando. Se mostró que los nuevos diseños encontrados son tan competitivos como el diseño óptimo local, que tiene un reducido número de puntos, con la ventaja que los diseños construidos tienen un mayor número de puntos experimentales. Se exhibió una familia de distribuciones de escala en la cual los diseños óptimos obtenidos son robustos a la elección de la distribución a priori, con respecto al cálculo de la eficiencia.

Referencias

- [1] H. Dette and W. Müller, “Optimal designs for regression models with a constant coefficient of variation,” *Journal of Statistical Theory and Practice*, vol. 7, no. 4, pp. 658–673, 2013. <https://doi.org/10.1080/15598608.2013.781833> 79
- [2] H. Dette and W. K. Wong, “Optimal Designs When the Variance Is A Function of the Mean,” *International Biometric Society*, vol. 55, pp. 925–929, 1999. <https://doi.org/10.1111/j.0006-341X.1999.00925> 79
- [3] D. Downing, V. V. Fedorov, and S. Leonov, “Extracting Information from the Variance Function: Optimal Design,” *mODa6 Physica, Heidelberg*, pp. 45–52, 2001. https://doi.org/10.1007/978-3-642-57576-1_6 79, 81
- [4] J. Gaviria and V. López-Ríos, “Locally d-optimal designs with heteroscedasticity: A comparison between two methodologies,” *Revista Colombiana de Estadística*, vol. 37, no. 1, pp. 93–108, 2014. 79, 82
- [5] J. Burrige and P. Sebastiani, “D-optimal designs for generalised linear models with variance proportional to the square of the mean,” *Biometrika*, vol. 81, no. 2, pp. 295–304, 1994. 79
- [6] A. C. Atkinson and R. D. Cook, “D-optimum designs for heteroscedastic linear models,” *Journal of the American Statistical Association*, vol. 90, pp. 204–212, 1995. 79
- [7] H. Dette and W. K. Wong, “Optimal bayesian designs for models with partially specified heteroscedastic structure,” *Annals of Statistics*, vol. 24, pp. 2108–2127, 1996. 79
- [8] Z. Fang and D. P. Wiens, “Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models,” *Journal of the American Statistical Association*, vol. 95, pp. 807–818, 2000. 79
- [9] B. P. M. Duarte and W. K. Wong, “Finding Bayesian Optimal Designs for Nonlinear Models: A Semidefinite Programming-Based Approach,” *International Statistical Review*, vol. 83, no. 2, pp. 239–262, 2015. <https://doi.org/10.1111/insr.12073> 79
- [10] H. T. Abebe, F. E. S. Tan, G. J. P. V. Breukelen, J. Serroyen, and M. P. F. Berger, “On the Choice of a Prior for Bayesian D-Optimal Designs for the Logistic Regression Model with a Single Predictor,” *Communications in Statistics - Simulation and Computation*, vol. 43, no. 7, pp. 1811–1824, 2014. <https://doi.org/10.1080/03610918.2012.745556> 79

- [11] I. Burghaus and H. Dette, “Optimal designs for nonlinear regression models with respect to non-informative priors,” *Journal of Statistical Planning and Inference*, vol. 154, pp. 12–25, 2014. <https://doi.org/10.1016/j.jspi.2014.05.009> 79
- [12] M. Buzoianu and J. B. Kadane, “Optimal Bayesian Design for Patient Selection in a Clinical Study,” *Biometrics*, vol. 65, no. 3, pp. 12–25, 2009. <https://doi.org/10.1111/j.1541-0420.2008.01156.x> 79
- [13] D. Braess and H. Dette, “On the number of support points of maximin and bayesian optimal designs,” *Ann. Statist.*, vol. 35, no. 2, pp. 772–792, 04 2007. <https://doi.org/10.1214/009053606000001307> 79
- [14] D. C. Woods, A. M. Overstall, M. Adamou, and T. W. Waite, “Bayesian design of experiments for generalized linear models and dimensional analysis with industrial and scientific application,” *Quality Engineering*, vol. 29, no. 1, pp. 91–103, 2017. <https://doi.org/10.1080/08982112.2016.1246045> 79
- [15] M. Goudarzi, “Bayesian optimal design for nonnegative binomial regression model with uniform prior,” *J Adv Pharm Edu Res*, vol. 9, no. S2, pp. 52–55, 2019. 79
- [16] O. A. Bodunwa, O. K. y Fasoranbaku, “D-optimal design in linear model with different heteroscedasticity structures,” *International Journal of Statistics and Probability*, vol. 9, no. 2, pp. 7–12, 2020. <https://doi.org/ijspv9n2p7> 79
- [17] V. I. Mosquera-Benitez, J. C. y López-Ríos, “Efecto de distribuciones a priori en los diseños d-óptimos bayesianos para un modelo no lineal correlacionado,” *Ciencia en Desarrollo*, vol. 10, no. 2, pp. 165–178, 2019. <https://doi.org/10.19053/01217488.v10.n2.2019.9504> 79
- [18] W. K. Duarte, B. P. M. y Wong, “Finding bayesian optimal designs for nonlinear models: A semidefinite programming-based approach,” *Int Stat Rev.*, vol. 83, no. 2, pp. 239–262, 2015. <https://doi.org/10.1111/insr.12073> 79
- [19] H. Abebe, F. Tan, G. Breukelen, and M. Berger, “Jmasm45: A computer program for bayesian d-optimal binary repeated measurements designs (matlab),” *Journal of Modern Applied Statistical Methods*, vol. 16, no. 1, pp. 689–721, 05 2017. <https://doi.org/10.22237/jmasm/1493599020> 79
- [20] C. Ritz and J. Streibig, *Nonlinear Regression with R*. Springer, New York, 2008. 80
- [21] V. I. López-Ríos and R. Ramos, “Una introducción a los diseños óptimos,” *Revista Colombiana de Estadística*, vol. 30, no. 1, pp. 37–51, 2007. 81

- [22] J. Kiefer and J. Wolfowitz, “The equivalence of two extremum problems,” *Canadian Journal of Mathematics*, vol. 12, no. 1, pp. 363–366, 1960. 83, 98
- [23] V. López-Ríos, *Diseños óptimos para discriminación y estimación en modelos no lineales*. Guanajuato, México: Centro de Investigación en Matemáticas A.C., 2008. 84
- [24] C. Patiño, “Diseños D_π -óptimos para modelos no lineales heteroscedásticos con inforación a priori,” Master’s thesis, Universidad Nacional de Colombia, 2017. 84
- [25] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. <http://www.R-project.org/> 86, 88, 89, 94, 96, 97
- [26] S. Rana, H. Midi, and A. Imon, “Robust wild bootstrap for stabilizing the variance of parameter estimates in heteroscedastic regression models in the presenceof outliers,” *Mathematical Problems in Engineering*, vol. 95, 2012. 93