Revista Ciencia en Desarrollo. Vol. 2 Num. 15. pp. 162- 169
julio - diciembre 2024

https://doi.org/10.19053/01217488.v15.n2.2024.15078

Artículo de revisión

Penalized Bayesian D-Optimal Designs for Regression Models of Continuous Response

Diseños D-Óptimos Bayesianos Penalizados para Modelos de Regresión de Respuesta Continua

Svetlana I. Rudnykh[1] ✉, Víctor Ignacio López-Ríos[2]

[1]Universidad del Atlántico, Barranquilla, Colombia.
[2]Universidad Nacional de Colombia, Medellín, Colombia. Escuela de Estadśtica.

Recepción: 02-oct-2022 **Aceptado:** 17-may-2024 **Publicado:** 23-jul-2024
Cómo citar: López-Ríos, V. I., & Rudnykh, S. I. (2024). Diseños D-Óptimos Bayesianos Penalizados para Modelos de Regresión de Respuesta Continua. Ciencia En Desarrollo, 15(2). https://doi.org/10.19053/01217488.v15.n2.2024.15078

Abstract

We propose extending the use of desirability functions in Bayesian optimal designs for regression models. This technique generates experimental designs with good statistical inference properties according to Bayesian optimal design theory and practical features, as defined by an investigator. These practical features are defined by a penalty function, using an overall desirability function, which is added to a Bayesian D-optimal design criterion to penalize impractical experimental designs. This methodology is illustrated by two examples of regression models: quadratic and exponential. Then, we compare designs obtained for different prior distributions of unknown parameters by efficiency calculations and simulation study. Results show that the D-efficiencies of the penalized designs relative to the non-penalized Bayesian D-optimal designs are competitive.

Keywords: Bayesian Optimal Designs; Desirability Functions; Exponential Growth Model; Penalized Designs; Quadratic Regression Model.

Resumen

Se propone extender el uso de funciones de deseabilidad en diseños óptimos bayesianos para modelos de regresión. Esta técnica genera diseños experimentales con buenas propiedades de inferencia estadística de acuerdo con la teoría del diseño óptimo bayesiano y características prácticas, según lo definido por un investigador. Estas características prácticas se definen mediante una función de penalización, utilizando una función de deseabilidad general, que se agrega a un criterio de diseño bayesiano D-óptimo para penalizar los diseños experimentales poco prácticos. Esta metodología se ilustra con dos ejemplos de modelos de regresión: cuadrático y exponencial. Luego, comparamos los diseños obtenidos para diferentes distribuciones previas de parámetros desconocidos mediante cálculos de eficiencia y estudios de simulación. Los resultados muestran que las D-eficiencias de los diseños penalizados en relación con los diseños D-óptimos bayesianos no penalizados son competitivas.

Palabras Clave: Diseños Óptimos Bayesianos; Funciones de Deseabilidad; Modelo de Crecimiento Exponencial; Diseños Penalizados; Modelo de Regresión Cuadrática.

E-ISSN: 2462-7658

1 Introduction

Experimental design plays a vital role in theoretical and applied scientific research. A well-designed experiment is an efficient method for learning about some phenomena, while a poorly designed experiment directly affects the quality of the conclusions derived from the experimental data. Statisticians have developed optimal design theory to generate efficiently designed experiments that satisfy the mentioned requirements.

For nonlinear models, optimal designs generally depend on the true values of the model parameters. Since the parameter vector is not known, the researcher must postulate a "best guess" of the unknown parameter vector resulting in locally optimal designs [1]. The problem may arise when that guess is not close enough to the true parameter vector, and therefore, the design obtained may not be optimal. In the Bayesian optimal design approach, the assumptions do not concentrate on single values. Instead, a prior distribution is assigned to each unknown parameter. These distributions can be centered around the assumed parameter values. The Bayesian optimality criterion is to minimize the Bayes risk by integrating the risk function over the prior distribution [2].

Furthermore, the optimal design theory can generate inadequate designs from a practical perspective. These designs can conflict with common practice in laboratories or other guidelines established. Many authors proposed several alternatives to generate optimal designs with the desired experimental properties, e.g. combination of several criteria [3]) and inclusion of restrictions or penalties in optimal design criteria [4], [5], [6].

A technique was proposed by [7] based on the desirability functions, as an alternative penalty approach. This technique helps to obtain optimal designs that fulfill practical design preferences.

Desirability functions have primarily been implemented in the manufacturing and industrial sectors. They are used by engineers to optimize product quality as the most popular and strongly suggested method to analyze several results simultaneously [8]. However, the desirability functions usually are not used in statistical procedures, particularly in the construction of optimal designs.

In this research, we suggest extending the use of desirability functions in Bayesian optimal designs for linear and nonlinear regression models. This procedure allows incorporating prior information of the unknown parameters by using a Bayesian approach and also satisfy practical preferences. A summary of the Bayesian *D*-optimality criterion for linear and nonlinear models is provided in Section 2. This summary is based on the work of [2]. A proposed penalized Bayesian *D*-optimal criterion for linear regression models is introduced in Section 3.1. Later this penalized Bayesian optimal design methodology is illustrated by an example of the quadratic regression model in Section 3.2. Furthermore, a proposed penalized Bayesian *D*-optimal criterion for nonlinear regression models is introduced in Section 3.3. Later this penalized design methodology for nonlinear models is illustrated by an example of the exponential growth model in Section 3.4. Discussion and conclusions are presented in Section 4.

2 Bayesian Optimal Designs

The idea of the Bayesian optimal design is to use any available prior information on the unknown parameters in the optimal design process. A Bayesian design problem is a problem of statistical decision [2], involving defining a design criterion, or a utility function $U(\xi, \boldsymbol{\theta}, \boldsymbol{y})$, which describes the worth (based on the experimental goals) of choosing the design ξ from the design space Ξ yielding data $\boldsymbol{y}$ from a sample space $\boldsymbol{Y}$, with model parameter values $\boldsymbol{\theta} \in \Theta$, where Θ is the parameter space. An (approximate) design ξ is a probability measure on the design space.

The expected gain in Shannon information is extensively used as an utility function ([2]). The Bayesian optimal design ξ^* maximizing the expected gain in Shannon information over the design space Ξ with respect to the future data $\boldsymbol{y}$ and model parameters $\boldsymbol{\theta}$ is the one that maximizes:

$$U_1(\xi) = \int \log \pi_1(\boldsymbol{\theta} \mid \boldsymbol{Y}, \xi)\, \pi_2(\boldsymbol{Y}, \boldsymbol{\theta} \mid \xi)\, d\boldsymbol{\theta}\, d\boldsymbol{Y}, \tag{1}$$

where $\pi_1(\boldsymbol{\theta} \mid \boldsymbol{Y}, \xi)$ and $\pi_2(\boldsymbol{Y}, \boldsymbol{\theta} \mid \xi)$ are posterior and join probability density functions, respectively. $U_1(\xi)$ represents the expected Shannon information of the posterior distribution.

2.1 Bayesian Optimality Criteria for Linear Models

Consider the problem of choosing a design ξ for a normal linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_N)^T$ is the vector of observations, $\boldsymbol{\varepsilon}$ is the $N \times 1$ vector of the errors and $\boldsymbol{X} = (\boldsymbol{f}(x_1), \boldsymbol{f}(x_2), \ldots, \boldsymbol{f}(x_N))^T$ is the $N \times p$ extended design matrix. $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, where $\boldsymbol{\theta} \in \Theta$, here Θ is an open convex set in $\mathbb{R}^p$. Under the model assumptions $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\theta}$ and $\mathrm{Cov}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I}_N$, where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix.

An approximate design ξ for this model is a probability measure on the design space X with finite support $x_1, \ldots, x_n$ and weights $w_1, \ldots, w_n$, representing the relative proportion of total observations taken at the corresponding design points.

Suppose that a prior distribution $\pi(\boldsymbol{\theta}, \sigma^2)$ on $\boldsymbol{\theta}, \sigma^2$ is given such that the conditional prior distribution $\pi(\boldsymbol{\theta}|\sigma^2)$ of $\boldsymbol{\theta}$ given σ^2 is $N(\mu, \sigma^2 \boldsymbol{R}^{-1})$, where $\boldsymbol{R}$ is a given positive definite $p \times p$ "precision" matrix. Under the mentioned assumptions the posterior conditional distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}, \sigma^2, \xi)$ of $\boldsymbol{\theta}$ given $\boldsymbol{Y}, \sigma^2$ is normal with mean vector

$$\widehat{\boldsymbol{\theta}}_B = E(\boldsymbol{\theta} \mid \boldsymbol{Y}, \sigma^2, \xi) = (N\boldsymbol{M}(\xi) + \boldsymbol{R})^{-1}(\boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{R}\mu) \tag{3}$$

and covariance matrix $\sigma^2(N\boldsymbol{M}(\xi) + \boldsymbol{R})^{-1}$, where $\boldsymbol{M}(\xi) = \frac{1}{N}\boldsymbol{X}^T\boldsymbol{X}$ is the Fisher information matrix for linear models (see details in [2] p. 277).

When Shannon information is considered as a utility function in the normal linear regression model, the design ξ^* maximizing the expected gain in Shannon information is the one that maximizes:

$$\begin{aligned} U_1(\xi) = &-\frac{p}{2}\log(2\pi) - \frac{p}{2} \\ &+ \frac{1}{2}\log\det\{\sigma^{-2}(N\boldsymbol{M}(\xi) + \boldsymbol{R})\}. \end{aligned} \tag{4}$$

After dropping the constant and multiplier terms in Equation (4), we can obtain the design optimality criterion

$$\Psi_1(\xi) = \det\left\{\boldsymbol{M}(\xi) + \frac{1}{N}\boldsymbol{R}\right\} = \det \boldsymbol{M}_B(\xi), \tag{5}$$

and it is known as Bayesian *D*-optimality for linear models, where non-Bayesian *D*-optimality maximizes the determinant of $\boldsymbol{M}_B(\xi)$. When the sample size N is large or the matrix $\boldsymbol{R}$ corresponds to imprecise information, the difference between a Bayesian design and its corresponding non-Bayesian one can be small; and the importance of the prior distribution disappears.

2.2 Bayesian Optimality Criteria for Nonlinear Models

Consider the nonlinear regression model

$$y_i = \eta(\mathbf{x}_i, \boldsymbol{\theta}) + \varepsilon_i, \tag{6}$$

where $\mathbf{x}$ is a $k \times 1$ vector of explanatory variables, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, when $\boldsymbol{\theta} \in \Omega$, here Ω is an open convex set in $\mathbb{R}^p$, $\eta(\mathbf{x}_i;\boldsymbol{\theta})$ is nonlinear in the model parameters, $Var(y_i)=\sigma^2$ and observations, y_i, are assumed to be independent.

Experimental design is usually more difficult to find in nonlinear models than in linear models. The reason is that their Fisher information matrix, $I(\xi,\boldsymbol{\theta})$, usually depends on the unknown parameters, which can not be separated as a simple multiplier.

$$\boldsymbol{I}(\xi,\boldsymbol{\theta}) = \int_X \left[\frac{\partial}{\partial\boldsymbol{\theta}}\eta(x;\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\boldsymbol{\theta}}\eta(x;\boldsymbol{\theta})\right]^T d\xi(x). \tag{7}$$

In non-Bayesian designs, the parameters in the Fisher information matrix are usually replaced by supposed values of the parameters, called "guesses" [1]. In the Bayesian optimal design approach, the assumptions do not concentrate on single values. Instead, a prior distribution is assigned to each unknown parameter. These distributions can be centered around the assumed parameter values. The Bayesian optimality criterion is to minimize the Bayes risk by integrating the risk function over the prior distribution [2].

The asymptotic approximations can be used for the nonlinear models since their exact posterior distributions are often intractable. Using the normal approximation in Equation (1), $U_1(\xi)$ can be written as

$$\begin{aligned} U_1(\xi) &= -\frac{p}{2}\log(2\pi) - \frac{p}{2} \\ &\quad + \frac{1}{2}\int \log\det\{N\boldsymbol{I}(\xi,\boldsymbol{\theta})\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}, \end{aligned} \tag{8}$$

where $\boldsymbol{I}(\xi,\boldsymbol{\theta})$ is the expected Fisher information matrix for a nonlinear model with unknown parameters $\boldsymbol{\theta}$ and N is a sample size.

The Bayesian optimality criterion can be obtained dropping the constant and multiplier terms in Equation (8)

$$\Psi_{BD}(\xi) = \int \log\det\{N\boldsymbol{I}(\xi,\boldsymbol{\theta})\}\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \tag{9}$$

It is known as Bayesian D-optimality criterion for nonlinear models, where Bayesian D-optimality maximizes the criterion $\Psi_{BD}(\xi)$.

3 Penalized Bayesian Optimal Designs

Experimental designs generated using optimal design theory may be inappropriate from a practical perspective. Those may conflict with common laboratory practice or other conventional guidelines. [7] proposed to combine an optimal design theory with desirability functions integrating desired experimental characteristics into optimal design.

[9] developed the concept of a desirability function to solve a multivariate optimization problem. This concept combines the responses of several factors into a single function in order to optimize the final outcome of a process.

Responses of each factor X_i, $i=1,2,\ldots,k$ are transformed to a dimensionless, ordinal measure d_i, $0\le d_i\le 1$, where a value of 0 designates the response as undesirable and a value of 1 indicates a desirable response. The intermediate values of the desirability scale can be consulted in [9]. The shape of the desirability function is determined by whether one is trying to maximize or minimize the response, or target a range of values, as shown in Figure 1. Each single desirability function can then be combined into composite desirability, which allows the simultaneous consideration of multiple constraints. The choice of appropriate desirability functions involved in obtaining penalized optimal designs with desirable characteristics is discussed in detail in papers [10] and [11].

The reasons for incorporating constraints (costs, or penalties) in optimal design may be different, depending on the purpose of the experiment. Also, the methods to incorporate the constraints may be different. The desirability functions transform desirable experimental properties to a dimensionless function with an ordinal scale. The use of desirability functions allows the user to obtain the experimental optimal design with important particular features. The advantage that these constraints may be described as continuous functions instead of fixed constraints [7].

In this work, we suggest extending the use of desirability functions in Bayesian optimal designs for linear and nonlinear models. Thus, the investigator can incorporate prior information of the unknown parameters by using a Bayesian approach and also satisfy practical preferences. This section presents our proposal to use them to obtain penalized Bayesian optimal designs.

3.1 Penalized Bayesian Optimality Criteria for Linear Models

We propose that a penalized Bayesian D-optimal design for linear models may be found by minimizing a new proposed criterion in:

$$\Psi_{PL}(\xi) = -\log\{\det\boldsymbol{M}_B(\xi)\} + \Lambda(1-D(\xi)) \tag{10}$$

with respect to $\xi\in\Xi$ for a given value of Λ, where $\Lambda>0$ is a user-specified scale constant. The first term of the penalized Bayesian optimal criterion in (10) represents a Bayesian D-optimality criterion for linear models (5), that is a monotone and convex function ([12]). The function $(1-D(\xi))$ in (10) is a bounded function between 0 and 1 ([9]), which is a penalty function representing constraints applied to the Bayesian D-optimal designs. The minimization of the criterion (10) is considered as the maximization of the expected utility (4), restricted by this penalty function. The penalized criterion (10) includes a user-defined parameter, Λ, that is required to control the balance between the overall desirability $D(\xi)$ (or penalty) and Bayesian optimality (5).

The Nelder-Mead direct search algorithm [13] is used to determine the penalized Bayesian optimal design ξ^* that minimizes the new criterion (10), given a value of Λ. The initial value of Λ is chosen by $\Lambda_0 = \left|\min_\xi\{-\log\{\det\boldsymbol{M}_B(\xi)\}\}\right|$, i.e., an absolute value of the minimum of the Bayesian D-optimality criterion (5). Penalized Bayesian optimal designs are generated by minimizing the penalized Bayesian optimal criterion (10) for values of Λ in submultiples of Λ_0. The final value of Λ is selected in the range where stability is exhibited in the responses of the overall desirability function $D(\xi)$. These responses can be plotted to better observe their behavior. The resulting optimal design is determined by the design support points and their corresponding weights of the penalized optimal design ξ_p^* associated with the minimum value of the penalized Bayesian optimal criterion for linear models (10) for a value of Λ in this range. The resulting penalized Bayesian optimal design is optimal according to the Bayesian D-optimal design criterion for linear models and the practical design preferences.

The methodology of the construction of the penalized Bayesian D-optimal design for linear models is illustrated with an example of the quadratic regression model.

3.2 Example 1: Quadratic Model

Consider the quadratic regression model ([12])

$$E(y) = \theta_0 + \theta_1 x + \theta_2 x^2 \tag{11}$$

and suppose $-1\le x\le 1$.

If the prior variances of $\theta_0,\theta_1,\theta_2$ are 3, 5 and 1 with absent correlation, and if sample size $N=9$, then the Bayesian D-optimal design is

$$\xi_B^* = \left\{\begin{array}{ccc} -1 & 0 & 1 \\ 0.369 & 0.261 & 0.369 \end{array}\right\}. \tag{12}$$

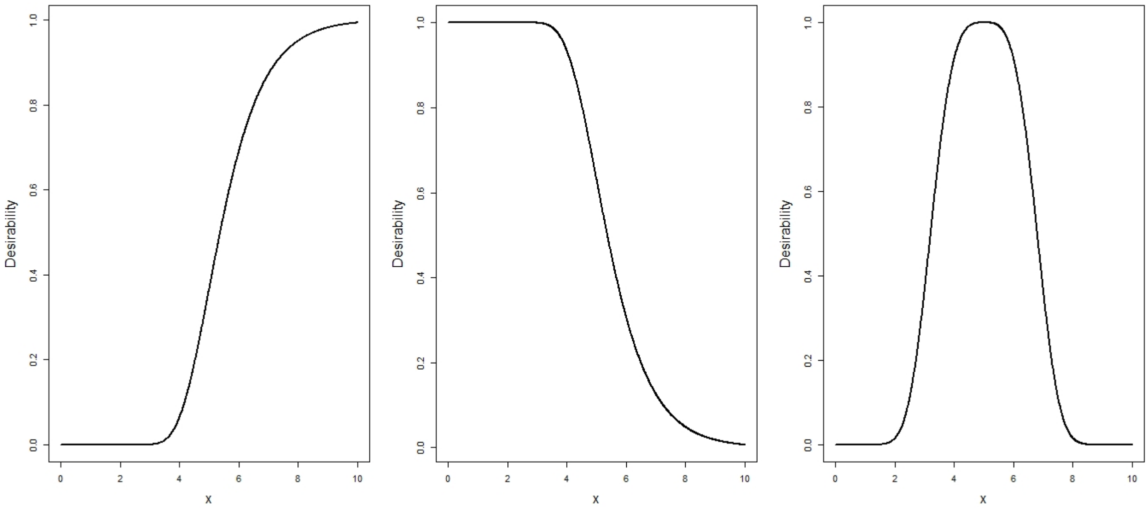

Figure 1: Bigger-is-better, smaller-is-better and target desirability functions.

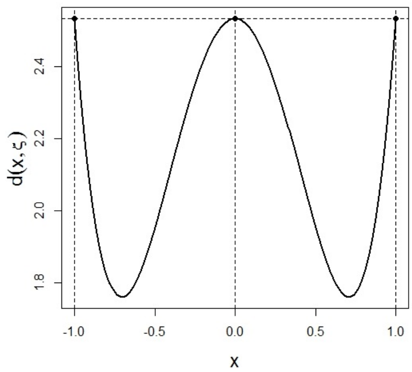

Figure 2: Example: quadratic model (11). Bayesian D-optimality verification for the three-point Bayesian optimal design (12).

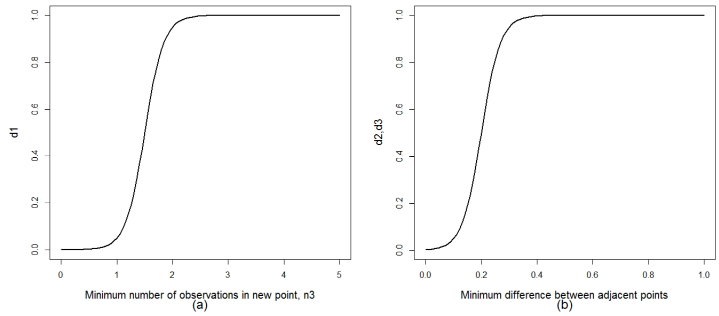

Figure 3: Plots of desirability functions for the three-point design (12).

Figure 2 shows that the function $d\,(x, \xi) = \boldsymbol{f}(x)^T \boldsymbol{M}_B(\xi)^{-1}\boldsymbol{f}(x)$ achieves its maximum value tr $\boldsymbol{M}(\xi^*)\boldsymbol{M}_B(\xi^*) = 2.5335$ at the design points $x = -1$, 0 and 1, demonstrating the D-optimality of the design ξ_B^* according to the equivalence theorem for Bayesian D-optimal designs for linear models ([14]).

The Bayesian design (12) contains three design support points. We want to have the four-point distinct D-optimal design with the minimum two observations in a new point placed between 0 and 1. In addition, the minimum difference between adjacent support points should be 0.3 units apart.

Initially, the fourth point is added to the initial design (12). A penalized-optimal design strategy was developed using three desirability functions to characterize the minimum number of observations in a new point x_3 (d_1) and the minimum difference between adjacent points (d_2 and d_3). A logistic cumulative distribution function, of the type described by [15], is used to create the desirability functions, though other functions can be used to achieve the appropriate shape. The logistic function, the form of the "bigger-is-better" or maximizing desirability function given in Figure 1(a), captures the experimental design preferences. The procedure of choosing the appropriate desirability functions with desirable characteristics can be consult in papers [10] and [11].

The desirability function d_1 to characterize the minimum number of observations n_3 in a new point x_3 is obtained as:

$$d_1\,(n_3) = \frac{1}{1 + \exp\left(-\left(n_3 - 1.5\right)/0.17\right)}. \tag{13}$$

A plot of this desirability function, given in Figure 3(a), shows that it is not acceptable to allocate less than one observation to the new support point x_3.

The desirability functions d_2 and d_3 to characterize the minimum difference (diff) between adjacent points x_2, x_3 and x_4 are obtained as:

$$d_2(\text{diff}_{23}) = d_3(\text{diff}_{34}) = \\ = \frac{1}{1 + \exp\left(-\left(\text{diff} - 0.2\right)/0.034\right)}. \tag{14}$$

A plot of the other desirability functions, given in Figure 3(b), shows that the spacing between support points of less than 0.1 units apart is unacceptable. Replacing the values of the design desired

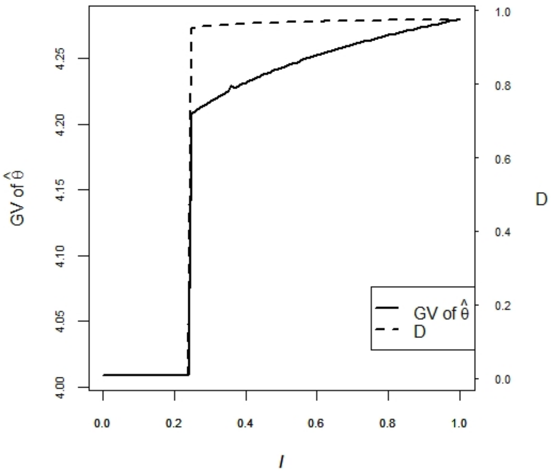

Figure 4: Tradeoff between overall desirability D and the generalized variance GV($\hat{\boldsymbol{\theta}}$) from the lth submultiple using the criterion given in (10).

characteristics in the formulas described by [15], the constants that appear in formulas (13) and (14) are obtained by [10] and [11].

The overall desirability function is

$$D = (d_1 \times d_2 \times d_3)^{1/3}. \qquad (15)$$

The Nelder-Mead direct search algorithm ([13]) is used in R-project ([16]) to minimize the penalized Bayesian optimal criterion in (10) using D and the quadratic model given in (11). The minimum Bayesian D-optimal criterion defined $\Lambda_0 = \left|\min_\xi \{-\log\{\det \boldsymbol{M}_B(\xi)\}\}\right| = 1.365$. Penalized Bayesian D-optimal designs were generated by minimizing the penalized Bayesian D-optimal criterion (10) for values of Λ in submultiples l of Λ_0 using the values $l = 0.01, 0.02, \ldots, 1$.

Initially penalized Bayesian D-optimal design is supported by only three design points. The fourth distinct design point appears when $l = 0.25$. Figure 4 graphically displays the responses of the generalized variance (GV) of $\hat{\boldsymbol{\theta}}$ and D of the penalized Bayesian optimal designs from the lth submultiple. It exhibits the balance between the imposed penalty for Bayesian designs without the desired characteristics and the generalized variance. Figure 4 indicates stability in the responses of the overall desirability function around $l = 0.4$. A penalized Bayesian D-optimal design from this range is

$$\xi^*_{BP} = \left\{ \begin{array}{cccc} -1 & 0 & 0.295 & 1 \\ 0.367 & 0.115 & 0.167 & 0.351 \\ 3 & 1 & 2 & 3 \end{array} \right\}, \qquad (16)$$

where the last row represents the number of observations at each design point for the sample size $N = 9$. The resulting design contains four distinct design points at least 0.3 units apart, and the new point placed between 0 and 1 has two observations.

Although this design has a small increase in the variance of the estimated parameters (4.232) as compared to Bayesian D-optimal design (4.008), this penalized optimal design has the practical characteristics desired by the researcher: four distinct support points with an acceptable number of observations assigned to each support point.

3.3 Penalized Bayesian Optimality Criteria for Nonlinear Models

The methodology proposed in Section 3.1, that combines the use of desirability functions and the Bayesian approach, can also be used in the construction of penalized Bayesian optimal designs for nonlinear regression models.

We suggest extending the use of desirability functions described in Bayesian optimal designs for nonlinear models. Thus, the researcher can incorporate prior information of the unknown parameters by using a Bayesian approach and also satisfy practical preferences. We propose that a penalized Bayesian D-optimal design (PBD) for nonlinear models may be found by minimizing with respect to $\xi \in \Xi$ for a given value of Λ the new criterion:

$$\Psi_{PBD}(\xi) = \int_\Theta -\log\det\{I(\boldsymbol{\theta}, \xi)\}\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} + \Lambda(1 - D(\xi)), \qquad (17)$$

where $\Lambda > 0$ is a user-specified scale constant.

The first term of the new criterion in (17) represents the Bayesian D-optimality criterion for nonlinear models, where $\Psi(\boldsymbol{\theta}, \xi) = -\log\det\{I(\boldsymbol{\theta}, \xi)\}$ is the D-optimality criterion for each $\boldsymbol{\theta} \in \Theta$. It follows from Jensen's inequality that if the functional $\Psi(\boldsymbol{\theta}, \xi)$ is convex then the Bayesian D-optimality criterion $\Psi_{BD}(\xi)$ (9) is also convex ([17]). The function $(1 - D(\xi))$ in (17) is a bounded function between 0 and 1 ([9]), which is a penalty function representing constraints applied to the Bayesian D-optimal designs. The minimization of the criterion (17) is considered as the maximization of the expected utility (8), restricted by the penalty function. This criterion is quasiconvex function by similar reasoning presented in [18, (p.113)], which determines the possibility of finding its global minimum. The penalized criterion (17) includes a user-deffined parameter, Λ, that is required to control the balance between the overall desirability $D(\xi)$ (or penalty) and Bayesian optimality (9).

The Nelder-Mead direct search algorithm ([13]) is used to determine the penalized Bayesian optimal design ξ^*_p that minimizes the new criterion (17), given a value of Λ. The methodology for the construction of penalized Bayesian D-optimal designs for nonlinear models is similar to that for linear models described in Section 3.1. The difference consists in the evaluation of the first term of the penalized Bayesian D-optimality criterion. To evaluate the integral in (17), the random variables $\boldsymbol{\theta}$ are generated according to respective prior distribution, and then the Monte Carlo method is used to calculate this integral. The resulting penalized Bayesian D-optimal design for a nonlinear model is optimal according to the Bayesian D-optimal design criterion (17) and also fulfills the practical design preferences.

The methodology for the construction of penalized Bayesian D-optimal designs for nonlinear models is illustrated with an example of the exponential growth regression model.

Table 1: Bayesian D-optimal designs for different values for α and β.

Values for α and β	$\alpha = \beta = 4$	$\alpha = \beta = 2.5$	$\alpha = \beta = 1.5$
Support points	0.0000; 0.9925	0.000; 0.937; 1.351	0.000; 0.675; 1.726; 6.431
Weights	0.5000; 0.5000	0.497; 0.373; 0.130	0.463; 0.307; 0.177; 0.052
Efficiency	0.99998	0.99964	0.96706

3.4 Example 2: Exponential Growth Model

Consider the exponential regression model with two parameters

$$\eta(x; \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 x), \quad x \geq 0,\ \theta_1 > 0,\ \theta_2 > 0, \qquad (18)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ denotes the unknown vector of parameters.

The non-Bayesian D-optimal designs for the exponential model (18) are balanced on exactly two support points for *all* values of the parameters $\boldsymbol{\theta}$ and do not depend on the parameter θ_1, i.e.,

$$\xi^* = \left\{ \begin{array}{cc} 0 & 1/\theta_2 \\ 0.5 & 0.5 \end{array} \right\}. \qquad (19)$$

The Bayesian D-optimal design for the model (18) depends on the prior distribution $\pi(\theta_1, \theta_2)$ only through the marginal distribution π_2 of θ_2. This design among all designs with two support points puts equal masses at the points

$$x_1 = 0, \quad x_2 = \left[\mathrm{E}_{\pi_2}(\theta_2)\right]^{-1}, \qquad (20)$$

where $\mathrm{E}_{\pi_2}[\cdot]$ denotes the expectation with respect to the marginal distribution π_2 of the prior π ([19]).

However, the Bayesian D-optimal designs for the exponential model (18) are not necessarily based on exactly two support points. The number of support points of Bayesian designs increases as the prior distribution for $\boldsymbol{\theta}$ becomes more dispersed ([20, 21]).

Since θ_2 is a scale parameter for the exponential model (18), it is admissible to adopt a gamma distribution for its prior ([19, 22]), i.e.

$$\pi_2(\theta_2) = \frac{\beta^\alpha \theta_2^{\alpha-1} \exp(-\beta\theta_2)}{\Gamma(\alpha)}, \quad \theta_2 > 0, \qquad (21)$$

where the hyperparameters α and β are positive and known. The hyperparameters α and β in the gamma prior distribution (21) are chosen such that $\mathrm{E}_{\pi_2}(\theta_2) = 1$, that is, $\alpha = \beta$ in all cases. Some of Bayesian D-optimal designs for different hyperparameters α and β are listed in Table 1.

The second row of this table shows the support points of the Bayesian D-optimal design, while the third row contains the corresponding weights. Figure 5 shows that the function $d\left(x, \xi_B^*\right)$ achieves its maximum value 2 at the support points, demonstrating the D-optimality of the resulting Bayesian designs according to the Equivalence Theorem for Bayesian D-optimal designs for nonlinear models [20].

The fourth row of this table presents the efficiencies of the non-Bayesian D-optimal design ξ^* given in (19) with respect to the Bayesian D-optimal designs given in Table 1. Here the efficiency is defined as the ratio of the Bayesian D-optimality criterion (9) evaluated for the non-Bayesian D-optimal design (19) and the Bayesian D-optimal designs. The efficiency values show that the Bayesian designs are competitive with respect to the non-Bayesian D-optimal design (19). A small improvement in the efficiency is observed when the Bayesian four-point design is used instead of the D-optimal two-point design.

This example shows that the number of support points of Bayesian design for the exponential model (18) is not fixed. When the prior distribution of θ_2 has support only over a small region, the Bayesian D-optimal designs have the same number of support points as non-Bayesian D-optimal design, and that the number of support points increases as the prior becomes more dispersed. A prior distribution with significant variance requires more support points for the Bayesian D-optimal design than a distribution with smaller variance. The same result is also obtained for the lognormal and uniform prior distributions ([23]).

It is noted that the Bayesian design for prior distribution of θ_2 with small variance contains two design support points. We want to have the three-point distinct Bayesian D-optimal design with the minimum two observations in new point and the minimum difference 0.3 units between adjacent design points for three prior distributions of θ_2: gamma, lognormal, and uniform.

Initially, the third point is added to the Bayesian D-optimal two-point design. A penalized-optimal design strategy is developed using two desirability functions $d_1(n_3)$ and $d_2(\mathrm{diff}_{23})$. A logistic cumulative distribution function, of the type described by [15], is used to generate the required desirability functions, but other functions can be used to obtain the appropriate shape. The desirability function $d_1(n_3)$ delimits the minimum number of observations in the new point x_3:

$$d_1(n_3) = \frac{1}{1 + \exp\left(-\left(n_3 - 1.5\right)/0.17\right)} \qquad (22)$$

and the desirability function $d_2(\mathrm{diff}_{23})$ defines the minimum difference between the design points x_2 and x_3:

$$d_2(\mathrm{diff}_{23}) = \frac{1}{1 + \exp\left(-\left(\mathrm{diff}_{23} - 0.2\right)/0.034\right)}. \qquad (23)$$

Replacing the values of the design desired characteristics in the formulas described by [15], the constants that appear in formulas (22) and (23) are obtained by [10]. Plots of these desirability functions can be found in Figure 6. The plot of $d_1(n_3)$, given in Figure 6(a), shows that the allocation of less than one observation to the new design point x_3 is unacceptable, and the plot of $d_2(\mathrm{diff}_{23})$, given in Figure 6(b), shows that the spacing between design points x_2 and x_3 of less than 0.1 units apart is unacceptable.

The overall desirability function is

$$D(\xi) = \left(d_1 \times d_2\right)^{1/2}. \qquad (24)$$

The Nelder-Mead direct search algorithm ([13]) is employed in R-project ([16]) to minimize the penalized criterion (17) for the exponential model given in (18). Penalized Bayesian D-optimal designs were generated by minimizing this criterion for values of Λ in submultiples l of Λ_0 using the values $l = 0.01, 0.02, \ldots, 1$, where Λ_0 is an absolute value of the minimum of the Bayesian D-optimality criterion (9). The overall desirability function D responses become stable from approximately $l = 0.10$. A resulting penalized Bayesian D-optimal design is obtained as

$$\xi_{BP}^* = \left\{ \begin{array}{ccc} 0.00 & 0.87 & 1.22 \\ 0.5 & 0.3 & 0.2 \end{array} \right\}. \qquad (25)$$

The same penalized-optimal design procedure is performed for lognormal and uniform prior distributions of θ_2, using the same two desirability functions $d_1(n_3)$ and $d_2(\mathrm{diff}_{23})$ defined in (22) and (23), respectively. The resulting penalized Bayesian D-optimal designs for these prior distributions coincide with the design (25) obtained for the gamma prior ([23]).

The efficiency of the design (25) with respect to the non-penalized Bayesian D-optimal two-point design is defined as the ratio of the Bayesian D-optimality criterion (9) evaluated for the non-penalized Bayesian design and corresponding penalized Bayesian design. It is observed that the D-efficiency of the penalized Bayesian design (25) relative to the non-penalized Bayesian D-optimal two-point designs is greater than of 99%, indicating the irrelevance of the loss in this efficiency. In summary, the penalized Bayesian D-optimal design (25) is as efficient as respective non-penalized design but it also has experimental characteristics desired by a researcher.

4 Discussion and Conclusions

We have proposed extending the use of desirability functions in Bayesian optimal designs to reduce issues related to experimental designs for linear and nonlinear models. A new optimality criterion was constructed with two simultaneous objectives: to incorporate the prior information of the unknown parameters and to satisfy practical design preferences imposed by a researcher. The proposed criterion combines the use of desirability functions and the Bayesian approach

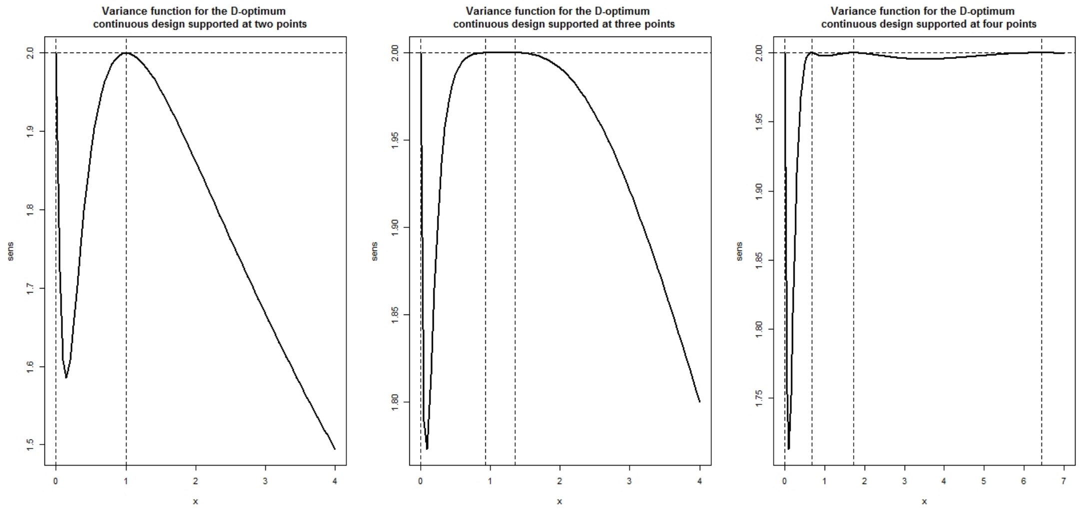

Figure 5: Expected variance $d\left(x, \xi_B^*\right)$ for the Bayesian D-optimal designs given in Table 1 for different values α and β.

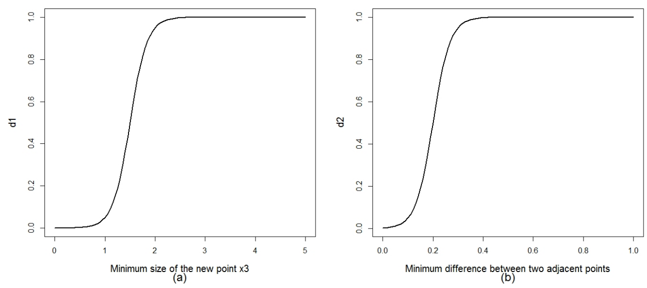

Figure 6: Plots of desirability functions for the exponential model (18).

in the construction of penalized Bayesian D-optimal designs, which have good statistical inference properties and desirable practical characteristics.

The proposed penalized optimal design criterion includes two terms, where the first term represents an "alphabetical" optimality criterion, and the second term is a penalty function that represents the constraints applied to impractical optimal designs. A parameter, Λ, specified by the researcher, manages the penalty and optimality contributions in the penalized optimal criterion. The initial value of Λ is suggested as an absolute value of the minimum of the corresponding non-penalized optimality criterion. It is recommended to choose the final value of Λ in the range corresponding to the stability exposed in the overall desirability function responses, which can be plotted to observe their behavior better. It is emphasized that because resulting penalized optimal designs are similar in the range of Λ that produces stable responses of the overall desirability, it is not important to know the exact final value of Λ.

The proposed penalized design procedure permits the researcher to define practical characteristics of the experimental design through individual desirability functions. The methodology of choosing the appropriate desirability functions according to the practical design preferences can be consulted in papers [10] and [11]. After defining the analytical expressions of these desirability functions, it is recommended to plot them to verify desirability levels of design characteristic subject to restriction; and then, it is suggested to use them in the proposed strategy. This allows for avoiding unpleasant errors when applying them in the penalty procedure.

This technique generates experimental designs with good statistical inference properties according to Bayesian optimal design theory and practical features, as defined by a researcher. These practical features are defined by a penalty function, using an overall desirability function, which is added to a Bayesian optimal design criterion to penalize impractical experimental designs. The resulting penalized Bayesian optimal design uses all available information of the unknown parameters and is statistically optimal in accordance with the Bayesian design criterion and the practical design preferences.

The proposed penalized Bayesian optimal design methodology was illustrated by two examples of regression models: quadratic and exponential. It was shown that the D-efficiencies of the penalized Bayesian D-optimal designs relative to the Bayesian D-optimal designs for different prior distributions were competitive, indicating the irrelevance of the loss in this efficiency.

Finally, the proposed methodologies allowed the construction of penalized Bayesian D-optimal designs that provide a suitable balance between Bayesian D-optimality and overall desirability. In all examples, this proposed penalized methodology has provided an experimental design that was optimal according to the Bayesian criteria and with the design preferences established by the researcher.

References

[1] H. Chernoff, "Locally Optimal Designs for Estimating Parameters", *The Annals of Mathematical Statistics*, vol. 24, no. 4, pp. 586-602, 1953.

[2] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review", *Statistical Science*, vol. 10, no. 3, pp. 273-304, 1995.

[3] G. E. Box and N. R. Draper, "Robust designs", *Biometrika*, vol. 62, no. 2, pp. 347-352, 1975.

[4] V. Dragalin and V. Fedorov, "Adaptive designs for dose-finding based on efficacy-toxicity response", *Journal of Statistical Planning and Inference*, vol. 136, no. 6, pp. 1800-1823, 2006.

[5] V. Dragalin, V. Fedorov and Y. Wu, "Adaptive designs for selecting drug combinations based on efficacy-toxicity response", *Journal of Statistical Planning and Inference*, vol. 138, no. 2, pp. 352-373, 2008.

[6] L. Pronzato, “Penalized optimal designs for dose-finding”, *Journal of Statistical Planning and Inference*, vol. 140, no. 1, pp. 283-296, 2010.

[7] S. M. Parker and C. Gennings, “Penalized locally optimal experimental designs for nonlinear models”, *Journal of agricultural, biological, and environmental statistics*, vol. 13, no. 3, pp. 334-354, 2008.

[8] F.-C. Wu, “Optimization of correlated multiple quality characteristics using desirability function”, *Quality Engineering*, vol. 17, no. 1, pp. 119-126, 2004.

[9] E. C. Harrington, “The desirability function”, *Industrial quality control*, vol. 21, no. 10, pp. 494-498, 1965.

[10] S. I. Rudnykh y V. I. López Ríos, “Elección de la función de deseabilidad para diseños óptimos bajo restricciones”, *Revista EIA*, vol. 15, no. 30, pp. 13-24, 2018.

[11] M. Shih, C. Gennings, V.M. Chinchilli and W.H Carter Jr, “Titrating and evaluating multidrug regimens within subjects”, *Statistics in medicine*, vol. 22, no. 14, pp. 2257-2279, 2003.

[12] A. DasGupta and W. Studden, “Robust bayesian experimental designs in normal linear models”, *The Annals of Statistics*, vol. 19, no. 3, pp. 1244-1256, 1991.

[13] J. A. Nelder and R. Mead, “A simplex method for function minimization”, *The computer journal*, vol. 7, no. 4, pp. 308-313, 1965.

[14] J. Pilz, “Bayesian estimation and experimental design in linear regression models”. John Wiley & Sons Inc, NY, 1991.

[15] R. D. Gibb, “Optimal treatment combination estimation for univariate and multivariate response surface applications”. Ph.D. thesis Virginia Commonwealth University, Richmond, Virginia, USA, 1998.

[16] R Core Team “R: A language and environment for statistical computing”. R foundation for statistical computing, Vienna, Austria, 2021, https://www.r-project.org/

[17] S. M. Ermakov and A. A. Zhiglijavsky, “Matematitscheskaja teorija optimalnich experimentov”. Nauka, Moskva, 1987.

[18] V. V. Fedorov and S. L. Leonov, “Optimal design for nonlinear response models”. CRC Press, 2013.

[19] H. Dette and S. Sperlich, “A Note on Bayesian D-optimal Designs for a Generalization of the Exponential Growth Model”, *South African Statistical Journal*, vol. 28, pp. 103-117, 1994.

[20] D. Firth and J. Hinde, “On Bayesian D-optimum Design Criteria and the Equivalence Theorem in Non-linear Models.”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 59, no. 4, pp. 793-797, 1997.

[21] B. P. 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.

[22] S. Mukhopadhyay and L. M. Haines, “Bayesian D-optimal Designs for the Exponential Growth Model.”, *Journal of Statistical Planning and Inference*, vol. 44, no. 3, pp. 385-397, 1995.

[23] S. I. Rudnykh, “Penalized bayesian optimal designs for nonlinear models of Continuous Response”. Ph.D. thesis Universidad Nacional de Colombia, Medellín, Colombia, 2020.