Journal of Statistical Modelling: Theory and Applications
Vol. 4, No. 1, 2013, pp. 59-73
Yazd University Press 2013

JSMTA

Research Paper

Nonparametric Bayesian optimal designs for unit exponential nonlinear model

Anita Abdollahi Nanvapisheh, Habib Jafari*, Soleiman Khazaei
Department of Statistics, Razi University, Kermanshah, Iran

Received: August 14, 2023/ Revised: December 30, 2023/ Accepted: January 22, 2024

Abstract: Nonlinear regression models have widespread applications across diverse scientific disciplines. Achieving precise fitting of the optimal nonlinear model is essential, taking into account the biases inherent in Bayesian optimal design. This study introduces a Bayesian optimal design utilizing the Dirichlet process as a prior. The Dirichlet process is a fundamental tool in exploring Nonparametric Bayesian inference, providing multiple well-suited representations. The research paper presents a novel one-parameter model, termed the "unit-exponential distribution", specifically designed for the unit interval. Additionally, a representation is employed to approximate the D-optimality criterion, considering the Dirichlet process as a functional tool. Through this approach, the aim is to identify a nonparametric Bayesian optimal design.

Keywords: Bayesian optimal design; D-optimal design; Dirichlet process; Nonparametric Bayesian; Stick-breaking prior; Unit exponential model.
Mathematics Subject Classification (2010): 62F15.

1 Introduction

In the field of experimental design, optimal design pertains to a specific category of designs categorized according to certain statistical criteria. It is widely recognized that a well-structured experiment can significantly improve the precision of statistical analyses. Essentially, before gathering data for model parameter estimation, it is advisable to select the most appropriate test that yields the maximum information. This selection, termed experimental design, is crucial, particularly when there is limited information available for the data and when measurements are costly in terms of

*Corresponding author: h.jafari@razi.ac.ir

time and expenses. As a result, many researchers have devoted their efforts to tackle the challenge of developing optimal designs for nonlinear regression models. Experimental design plays a crucial role in various scientific research domains, including but not limited to biomedicine and pharmacokinetics. Its application in these fields allows researchers to conduct rigorous investigations and obtain valuable insights.

Optimal designs are pursued by employing optimality criteria, typically grounded in the information matrix. Until 1959, research primarily concentrated on linear models, where the models exhibited linearity concerning the parameters. However, in nonlinear models, the introduction of unknown parameters added intricacies to the design problem, given that the optimality criteria relied on these unknown parameters (Atkinson et al., 2007; Bürkner et al., 2019). To tackle this challenge, researchers proposed various solutions, encompassing local optimal designs (Aminnejad and Jafari, 2017; Chernoff, 1953; Dette et al., 2006; Ford and et al., 1992; Rodríguez-Torreblanca and Rodríguez-Díaz, 2007), sequential optimal designs, minimax optimal designs, Bayesian optimal designs (Parsa Maram and Jafari, 2016; Goudarzi and et al., 2019; Graßhoff and et al., 2012; Kiefer, 1959; Kiefer and Wolfowitz, 1959), and pseudo-Bayesian designs (Mukhopadhyay and Haines, 1995). Chernoff (1953) introduced the concept of local optimality, which entails specifying fixed values for the unknown parameters and optimizing a function of the information matrix to determine the design for these specified parameter values.

The determination of unknown parameter values in local designs is typically derived from prior studies or experiments conducted explicitly for this purpose. The efficacy of local designs heavily hinges on the judicious selection of these parameter values. However, a notable challenge arises when the investigated problem lacks robustness concerning weak parameter estimation. To address this, an alternative approach for local optimal designs involves incorporating a prior distribution for the unknown parameters rather than relying solely on an initial guess. In the Bayesian method, the initial step is to encapsulate the available information in the form of a probability distribution for the model parameter, known as the prior distribution. A Bayesian optimal design seeks to maximize the relevant optimality criterion over this prior distribution. Nevertheless, it is crucial to acknowledge that the choice of the prior distribution within the Bayesian framework can be problematic and may potentially yield erroneous results. The selection of the prior distribution is subjective, relying on the researcher's beliefs, and it could influence the final outcome. Unfortunately, the Bayesian approach lacks a definitive method for selecting the prior distribution. Numerous researchers have delved into investigating the impact of the prior distribution on determining design points in various types of optimal designs. For instance, Chaloner and Larntz (1989); Chaloner and Duncan (1983); Burghaus and Dette (2014); Chaloner and Verdinelli (1955); Pronzato and Walter (1985); Mukhopadhyay and Haines (1995); Dette and Neugebauer (1996, 1997); Fedorov and Hackl (2012); Fedorov and Leonov (2013); Firth and Hinde (1997) have made significant contributions to this field. For additional insights into this subject, Chapter 18 of Atkinson et al.'s book (Atkinson et al., 2007) serves as a valuable resource. Furthermore, in situations where there is insufficient evidence from prior studies on the relevant topic, specifying an appropriate prior distribution becomes a challenging task. In such instances, subjective or noninformative prior distributions are employed, encompassing all available information pertaining to the uncertainty of

the parameter values. For further details, refer to Burghaus and Dette (2014).

This research paper presents a unique one-parameter distribution known as the unit-exponential (UE) distribution, explicitly tailored for the unit interval in Section 2. Subsequently, the introduced distribution is treated as a nonlinear regression model, and its optimal designs are elucidated. To achieve this, Section 3 commences with discussions on optimal designs for nonlinear models, followed by the derivation of pertinent relationships for our model. The paper concludes in Section 4, offering closing remarks.

2 The unit-exponential distribution

If Y follows the exponential distribution with the probability density function (pdf) $g(y) = \theta \exp(-\theta y)$, then employing the transformation $X = \frac{Y}{1+Y}$ yields a new distribution with support on the unit interval, where the cumulative distribution function (cdf) and the pdf of the resultant distribution are given by

$$F(x \mid \theta) = 1 - \exp(\frac{-\theta x}{1-x}), \qquad 0 < x < 1, \quad \theta > 0, \tag{1}$$

$$f(x \mid \theta) = \frac{\theta}{(1-x)^2} \exp(\frac{-\theta x}{1-x}), \qquad 0 < x < 1, \quad \theta > 0. \tag{2}$$

The hazard rate function (hrf) for this distribution is expressed as follows

$$h(x \mid \theta) = \frac{f(x \mid \theta)}{1 - F(x \mid \theta)} = \frac{\theta}{(1-x)^2}, \qquad 0 < x < 1, \quad \theta > 0.$$

In Figure 1, the pdf and the hrf of this distribution are graphed for various values of the parameter θ. According to this figure, it is evident that the hrf exhibits an increasing trend within the interval $0 < x < 1$. Moving forward, we will explore several key statistical properties of the UE distribution.

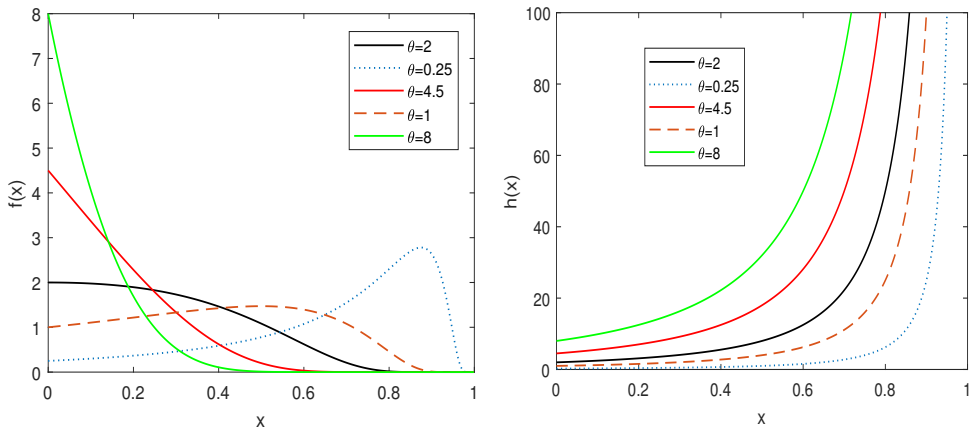

Figure 1: Plot of pdf (left) and hrf (right).

Next, we shall discuss several main statistical properties of the UE distribution.

2.1 Moments

The k-th moment about the origin of the UE distribution is provided by the expression (Bakouch et al., 2023):

$$\mu_k' = E(X^K) = K \int_0^1 x^{k-1} \exp(\frac{-\theta x}{1-x}) dx, \qquad k = 1, 2, \ldots,$$

which can not be solved analytically. Various moments of UE distribution have been shown in Table 1, for $k = 1, 2, 3, 4$ and solved using Maple software.

Table 1: Various Moments of UE distribution.

	$\theta = 0.2$	$\theta = 0.5$	$\theta = 1$	$\theta = 1.5$	$\theta = 2$	$\theta = 2.5$
μ_1'	0.7013	0.5385	0.4036	0.3276	0.2773	0.2411
μ_2'	0.5429	0.3463	0.2109	0.1466	0.1093	0.0853
μ_3'	0.4388	0.2407	0.1237	0.0756	0.0507	0.0361
μ_4'	0.3639	0.1753	0.0777	0.0425	0.0260	0.0170
SD	0.2260	0.2372	0.2190	0.1979	0.1800	0.1648
CV	0.3222	0.4404	0.5426	0.6040	0.6491	0.6835
Sk	−1.1762	−0.4816	−0.0162	0.2362	0.4148	0.5431
K	3.5924	2.2348	1.9617	2.1261	2.3359	2.4364

In this table, SD is the Standard Deviation and CV is the coefficient of variation of new distribution.The skewness (Sk) and kurtosis (K) of the distribution are given by

$$\begin{aligned} \text{skewness}(Sk) &= \frac{\mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3}{(\mu_2' - \mu_1'^2)^3/2}, \\ \text{kurtosis}(K) &= \frac{\mu_4' - 4\mu_1'\mu_3' + 6\mu_1'^2\mu_2' - 3\mu_1'^4}{(\mu_2' - \mu_1'^2)^2}. \end{aligned}$$

Higher-order moments can be computed numerically through the use of software such as MAPLE. As indicated in Table 1, the moment decreases with an increase in θ, while skewness exhibits an increasing trend with rising θ. On the other hand, the kurtosis does not demonstrate a specific trend.

2.2 Incomplete moments

The *kth* incomplete moment of the UE distribution is expressed as per the formula provided by Mazucheli et al. (2019)

$$T_k(t) = E[X^k | x < t] = \int_0^t x^k \exp(\frac{-\theta x}{1-x}) dx, \qquad k = 1, 2, \ldots,$$

which can not be solved analytically. In particular, for $k = 1, 2$ and by considering the change of variables, $u = (\frac{\theta x}{1-x})$, we obtain

$$T_1(t) = \exp(\theta)(-(\exp(-(\theta + \frac{\theta t}{(1-t)})) - \exp(-\theta)) + \theta(Ei(1, \frac{\theta t}{(1-t)}) - Ei(1, \theta))),$$

where

$$Ei(a,z) = \int_1^{\infty} z^{-a} \exp(-yz) dy.$$

Similarly, the $T_2(t)$ is obtained.

2.3 Mean deviation

As pointed out, like study of Ghitany et al. (2008), the amount of scattering in a population is measured to some extent by the totality of deviations from the mean and the median. These are known as the mean deviation about the mean and the mean deviation about the median, defined as (Mazucheli et al., 2019)

$$\delta_1(X) = \int_0^{\infty} |x-\mu| f(x) dx = 2\mu F(\mu) - 2I(\mu),$$

that

$$\begin{aligned} I(\mu) &= \int_0^{\mu} x f(x) dx, \\ \delta_2(X) &= \int_0^{\infty} |x-M| f(x) dx = \mu - 2I(M), \end{aligned}$$

where $\mu = E(X)$ and $M = Median(X)$. So we have

$$\begin{aligned} \delta_1(X) &= 2\mu F(\mu) - 2I(\mu) = 2\mu(1 - \exp(\frac{-\theta\mu}{1-\mu})) - 2\int_0^{\mu} x f(x) dx, \\ \delta_2(X) &= \mu - 2I(M) = \mu - 2\int_0^{M} x f(x) dx, \end{aligned}$$

2.4 Quantile function

Let X be a unit-exponential random variable with cdf (1). The quantile function $Q(p) = F^{-1}(p)$ can be written as $1 - \exp(\frac{-\theta Q}{1-Q}) = p$, then we have $Q(p \mid \theta) = \dfrac{\log(1-p)}{\log(1-p) - \theta}$.

2.5 Maximum likelihood estimator

In this section, we will explore the estimation of the parameter θ for the UE distribution using both maximum likelihood methodology and the method of moments. Suppose $X_1, \ldots, X_n$ form a random sample from the UE distribution with the pdf given by (2). The log-likelihood function of θ can then be expressed as:

$$\ell(\theta \mid \boldsymbol{x}) \propto n \log\theta + \sum_{i=1}^{n} \log \frac{1}{(1-x_i)^2} - \theta \sum_{i=1}^{n} \frac{x_i}{(1-x_i)}. \tag{3}$$

The maximum likelihood estimate of θ is derived by taking the derivative of (3) with respect to the parameter. This yields $\hat{\theta} = \dfrac{n}{u(\boldsymbol{x})}$, where $u(\boldsymbol{x}) = \sum\limits_{i=1}^{n} \dfrac{x_i}{(1-x_i)}$.

2.6 Simulation study

In this section, a Monte Carlo simulation is conducted to assess and compare the finite sample behavior of the maximum likelihood estimators for the parameterθ. Samples of sizes $n = 50, 100, 150$ and 200 are generated, considering various values of the parameter θ. To simulate observations from the UE distribution, Y is generated from an exponential distribution, and then the transformation $X = Y/(1 + Y)$ is applied. The simulation experiment is repeated $N = 1000$ times. The results for bias and Mean Squared Error (MSE) are presented in Table 2 indicate a positive bias in the MLE of the parameter θ The results are consistently stable and, notably, closely approximate the actual values for this sample size. Furthermore, the MSE decreases with an increase in the sample size.

Table 2: Estimated bias (mean-squared error) of θ by maximum likelihood method.

θ	n	Bias	MSE	θ	n	Bias	MSE
0.1	50	0.0944	0.7092	0.5	50	0.5502	0.4258
	100	0.0362	0.0753	0.5	100	0.4779	0.2737
	150	0.0155	0.0242	0.5	150	0.4767	0.2607
	200	0.0139	0.0005	0.5	200	0.4556	0.1374
2	50	0.9485	1.0271	5	50	0.5863	0.6224
	100	0.8950	0.9555	5	100	0.4167	0.3036
	150	0.8835	0.8453	5	150	0.3466	0.2020
	200	0.8680	0.8022	5	200	0.2902	0.1570

2.7 Applications

In this section, we outline the applications of the newly proposed distribution and demonstrate that the UE distribution offers a superior fit compared to other models for the given dataset. All computations are executed using R and MATLAB softwares. The example focuses on the monthly water capacity data sourced from the Shasta reservoir in California, USA, spanning from February 1991 to 2010, as retrieved from http://cdec.water.ca.gov/reservoirmap.html, in line with Mukhopadhyay and Haines (1995). A crucial criterion for determining the appropriateness of a particular distribution for a dataset is the empirical hazard function of the data. In this context, we employ the scaled Total Time on Test (TTT) function to identify the type of hrf exhibited by the data and subsequently select the most suitable distribution (Aarset, 1987). The TTT plot for the dataset is depicted in Figure 2. Finally, to distinguish between the unit exponential, unit-Rayleigh, unit-Weibull, and unit-BXII distribution, we compute $-2\log(L)$, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). The results are summarized in Table 3. The optimal model corresponds to lower $-2\log(L)$, and BIC values. The values of these measures suggest that the UE distribution is a robust contender against other competitive distributions and, moreover, stands out as the best fit among them.

Figure 3 illustrates the histogram of the dataset alongside the fitted density function and plots of the empirical and estimated pdf for these fitted distributions. The figure provides evidence supporting the conclusion that the UE distribution is a suitable model for fitting the considered dataset.

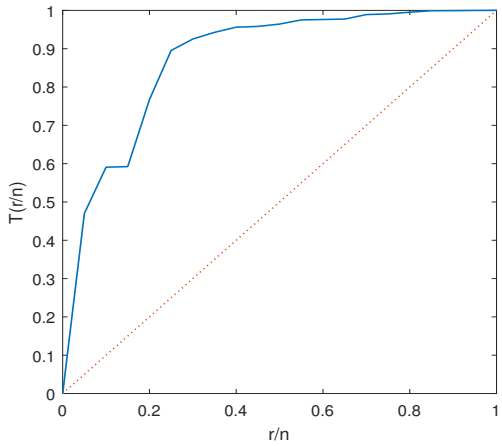

Figure 2: TTT plot for the data set.

Table 3: Parameter estimates, log-likelihood values and goodness-of-fit measures for the data set.

model	α	θ	β	$-2\log(L)$	AIC	BIC
Unit-exponential	–	0.29710	–	−22.5453	−20.5453	−19.5495
Unit-Rayleigh	–	–	5.215272	−17.23051	−15.2305	−12.2347
Unit-Weibull	4.207084	–	1.57037	−21.9138	−17.9138	−15.9224
Unit-BXII	5.856231	–	1.76528	−23.4884	−19.4884	−17.4969

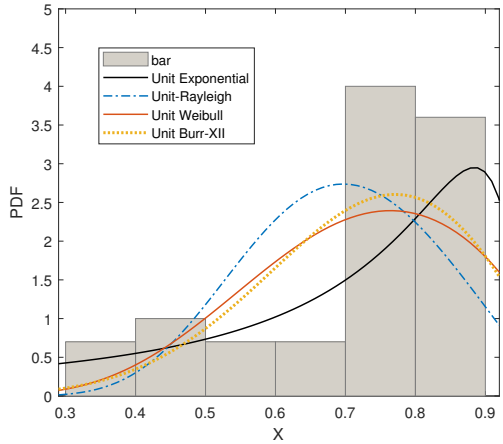

Figure 3: The plots of the fitted pdfs of the considered distributions as well as the histograms for the data set.

3 Optimal design for nonlinear models

In the realm of nonlinear experimental design, a common scenario arises where in the relationship between the response variable y and the independent variable x is given by

the equation $y = \eta(x, \boldsymbol{\theta}) + \epsilon$ where $x \in \chi \subseteq R$ and y is a response variable and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is the unknown parameter vector and ϵ is a normally distributed residual value with mean 0 and known variance $\sigma^2 > 0$. For simplicity, we assume $\sigma^2 = 1$ in this problem. If $\eta(x, \boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$ then, the information matrix $M(\xi, \boldsymbol{\theta})$ at a given point x can be represented as follows (Goudarzi and et al., 2019)

$$I(\xi, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \eta(x, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}^T} \eta(x, \boldsymbol{\theta}).$$

Several optimality criteria are employed to attain the optimal design, including D-optimality and A-optimality. These criteria are functions of the information matrix and can be expressed as follows

$$\Psi_D(\xi, \boldsymbol{\theta}) = -\log(\det(M(\xi, \boldsymbol{\theta}))), \Psi_A(\xi, \boldsymbol{\theta}) = \ tr(M^{-1}(\xi; \boldsymbol{\theta})),$$

where ξ denotes a design with two components; the first component represents specific values from the design space χ and the second component corresponds to the weights assigned to these values, so that design ξ can be defined as follows

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_\ell \\ w_1 & w_2 & \dots & w_\ell \end{array} \right\} \in \boldsymbol{\Xi},$$

where $\boldsymbol{\Xi} = \{\xi \mid 0 \leq w_j \leq 1; \sum_{j=1}^{\ell} w_j = 1, x \in \boldsymbol{\chi}\}$, (Kiefer, 1974).

When considering a discrete probability measure ξ with finite support, the information function of ξ can be expressed as follows (Atkinson et al., 2007)

$$M(\xi, \boldsymbol{\theta}) = \sum_{j=1}^{\ell} w_j I(x_j, \boldsymbol{\theta}).$$

Because of the dependence of the information matrix $M(\xi, \boldsymbol{\theta})$ to the unknown parameter $\boldsymbol{\theta}$, one approach to address this issue is to employ the Bayesian method and incorporate a prior distribution of the parameter vector. The Bayesian D-optimality criterion can be formulated as follows

$$\Psi_\Pi(\xi) = E(\psi(\xi; \boldsymbol{\theta})) = \int_\Theta \psi(\xi; \boldsymbol{\theta}) d\Pi(\boldsymbol{\theta}) = \int_\Theta -\log(\det(M(\xi, \boldsymbol{\theta}))) d\Pi(\boldsymbol{\theta}), \qquad (4)$$

where Π represents the prior distribution for $\boldsymbol{\theta}$ and the Bayesian D-optimal design is attained by minimizing (4). According to Dette and Neugebauer (1996), in the general case of optimal designs which can include designs with two and more points, if the support of the prior distribution has n points, then the maximum number of Bayesian optimal design points is determined by $n^{\frac{p(p+1)}{2}}$, where p is the model parameter. Hence, in the specific scenario of nonlinear models with one parameter ($p = 1$), this implies that the support of the Bayesian optimal design does not contain more points than the support of the prior distribution.

In certain situations, it can be challenging for the experimenter to specify a prior distribution on the $\boldsymbol{\Theta}$ parameter space. In such cases, an alternative approach is to consider an unknown prior distribution Π for the parameter $\boldsymbol{\theta}$. In this condition, Π

is treated as a parameter itself. Consequently, (4) becomes a random function, and it becomes necessary to determine its distribution or approximation. From a Bayesian perspective, we construct a prior distribution on the space of all distribution functions to address this issue. Ferguson (1973) introduced the concept of the Dirichlet process (DP) in this context, that in the section 3.1.1 an overview of the DP will be provided in section 3.1.1.

3.1 Nonparametric Bayesian D-optimal design

In this section, we introduce the nonparametric Bayesian optimal design. In the nonparametric Bayesian framework, it is assumed that $\theta \mid P \sim P$, where P is a random probability distribution and $P \sim \Pi$. The general method of construction a random measure involves starting with the stochastic processes. Ferguson (1973) formulated the requirements that must be imposed on a prior distribution and proposed a class of prior distributions, named DPs. One of the main argument for using the Dirichlet distribution in practical applications is based on the fact that this distribution serves as good approximation of many parametric probability distributions. Below we provide the definition of the DP.

3.1.1 Dirichlet process

To establish a random distribution G distributed according to a DP, its marginal distributions must follow a Dirichlet distribution. Specifically, let H be a distribution over $\mathbf{\Theta}$ and α be a positive real number. For any finite measurable partition $A_1, \ldots, A_r$ of $\mathbf{\Theta}$, the vector $(G(A_1), \ldots, G(A_r))$ is random since G is random. We say G is a DP distributed with base distribution H and concentration parameter α, written $G \sim \mathrm{DP}(\alpha, H)$ (Ferguson, 1973), if the following conditions hold

$$(G(A_1), \ldots, G(A_r)) \sim Dir(\alpha H(A_1), \ldots, \alpha H(A_r)), \tag{5}$$

for every finite measurable partition $A_1, \ldots, A_r$ of $\mathbf{\Theta}$.

The parameters H and α play intuitive roles in the definition of the DP. The base distribution H represents the mean of the DP, such that for any measurable set $A \subset \mathbf{\Theta}$, we have $E[G(A)] = H(A)$. On the other hand, the concentration parameter α can be viewed as an inverse variance: $V[G(A)] = H(A)(1 - H(A))/(\alpha + 1)$. The larger α is, the smaller the variance, and the DP will concentrate more of its mass around the mean. The concentration parameter is also referred the strength parameter, referring to the strength of the prior when using the DP as a nonparametric prior in Bayesian nonparametric models. It can be interpreted as the amount of mass or sample size associated with the observations. It is worth noting that α and H only appear as their product in the definition of the DP (5). Consequently, some authors treat $H = \alpha H$, as the same as the single (positive measure) parameter of the DP, writing $\mathrm{DP}(H)$ instead of $\mathrm{DP}(\alpha, H)$. This parameterization can be notationally convenient, but loses the distinct roles α and H play in describing the DP.

Consequently, when α approaches infinity ($\alpha \to \infty$), $G(A)$ approaches $H(A)$ for any measurable set A, indicating weak or pointwise convergence of G to H. However, it is important to note that this does not imply a direct convergence of G to H as a whole. In fact, as we will explore later, samples drawn from a DP will typically be discrete

distributions with probability one, even if the base distribution H is smooth. Therefore, G and H may not be absolutely continuous with respect to each other. Despite this, some authors still utilize the DP as a nonparametric extension of a parametric model represented by H. However, if the desire is to maintain smoothness, it is possible to extend the DP by convolving G with kernels, resulting in a random distribution with a density function.

An alternative definition of the DP is proposed by Ferguson (1973) that defined a random probability measure which is a DP on $(\boldsymbol{\Theta}, B(\boldsymbol{\Theta}))$, as

$$P(.) = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}(.),$$

where $\theta_i (i > 1)$ is a sequence of i.i.d. random variables with common distribution Q, δ_{θ_i} represents a probability measure that is degenerate at θ that $\delta_{\theta_i} = 1$ if $\theta_i \in A$ and 0 otherwise, and p_i's are the random weights satisfying $p_i \neq 0$ and $\sum_{i=1}^{\infty} p_i = 1$. The random distribution P is discrete with probability one. Several authors have proposed alternative series representations of the DP. Bondesson (1982); Sethuraman (1994); Zarepour and Al Labadi (2012) are among those who have contributed to this area. In the upcoming section, we will discuss the nonparametric Bayesian D-optimal design for the UE model.

3.1.2 Nonparametric Bayesian D-optimal design for UE model

Now, let's consider the following regression model

$$E(y|x) = \eta(x, \boldsymbol{\theta}) = \frac{\theta}{(1-x)^2} \exp(\frac{-\theta x}{1-x}), \qquad 0 < x < 1, \quad \theta > 0. \tag{6}$$

Therefore, the Bayesian D-optimality criterion, denoted as $\Psi_{\Pi}(\xi)$ can be expressed according to our model as follows (Goudarzi and et al., 2019)

$$\Psi_{\Pi}(\xi) = E(\psi(\xi; \boldsymbol{\theta})) = \int_{\Theta} -\log(\sum_{j=1}^{\ell} w_j [\exp(\frac{-\theta x_j}{1-x_j})(\frac{1}{(1-x_j)^2} + \frac{\theta x_j}{(1-x_j)^3})]^2) d\Pi(\theta), \tag{7}$$

where Π is the prior distribution for θ. The Bayesian D-optimal design is attained by minimizing (7). In the nonparametric Bayesian framework, we consider $P \sim \mathrm{DP}(\alpha, P_0)$ and its collective representation as $P(.) = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}(.)$. In this context, the optimality criterion can be expressed as follows

$$\Psi_{\Pi}(\xi) = \sum_{i=1}^{\infty} p_i (-\log(\sum_{j=1}^{\ell} w_j [\exp(\frac{-\theta_i x_j}{1-x_j})(\frac{1}{(1-x_j)^2} + \frac{\theta_i x_j}{(1-x_j)^3})]^2)). \tag{8}$$

Chernoff (1953) demonstrated that when searching for a local optimal design, there exists an optimal design where all the mass is concentrated at a single point within the design supports. Caratheodory's theorem (Atkinson et al., 2007) also confirms the

existence of a one-point optimal design. However, when employing the Bayesian optimality criterion, a more complex situation arises. Brice (2006) showed that with a uniform prior distribution, as the support of the prior distribution increases, the number of optimal design points for the single-parameter model also increases. Challoner (1983) suggested that if the researcher seeks to achieve a one-point optimal design, it is advisable to consider a small support for the uniform prior distribution. The same principle applies to nonparametric Bayesian designs. In this scenario, assuming a uniform distribution over the interval $[1, B]$ (we considered the values of 10, 50, 100 and 300 for B in this study) as the basic distribution, the one-point optimal design can be achieved.

Equation (7) represents a stochastic function of the DP. According to Ferguson's definition of the DP, the calculation of (8) is not straightforward directly. To address this challenge and obtain an approximation of the optimal nonparametric Bayesian criterion, methods such as the stick-breaking process is employed. Sethuraman (1994) introduced this method as a significant approach for generating realizations of the DP, that we used this method.

In this section, we employ a uniform distribution as the base measure in the DP and obtain the results by using a nonlinear optimization programing R package Rsolnp. We consider uniform distribution on the interval $[1, B]$ as the base measure in DP, that is, uniform distribution is assumed as an initial guess for distribution of θ. To better understanding of the effect of the α parameter, we tabulate the results for four different values of $\alpha = 1, 5, 10, 50$, in Table 4. We also fixed $\epsilon = 10^{-10}$. Without loss of generality, we consider a bounded design space $\chi = [0, 1]$.

Table 4 represents the results when the concentration parameter and uncertainty in the base measure increase.

According to the results, as the value of α increases, the support points in two-point design do not change significantly. The weight of the minimum point increases rapidly and the smallest point will have the most weight that this weight almost increases or remains fixed by increasing the concentration parameter. In addition, in the range under investigation, the results show that we do not have a three-point design for all parameter spaces $[1, B]$ and in fact, it converts to the design by less points.

4 Discussion and conclusions

Nonlinear regression models are widely used in various scientific fields and the Bayesian method is commonly employed to obtain optimal designs in such models. However, one of the challenges in the Bayesian framework is the subjective selection of the prior distribution, which can potentially lead to incorrect results. The choice of the prior distribution is often based on the researcher's beliefs, and it strongly influences the final outcome. Unfortunately, the Bayesian approach lacks a systematic method for selecting the prior distribution. To overcome these limitations and reduce reliance on restrictive parametric assumptions, nonparametric Bayesian methods are pursued.

In this study, we consider the prior distribution as an unknown parameter and utilize the DP to derive nonparametric Bayesian D-optimal designs. Specifically, we focus on a nonlinear model with one parameter, namely the unit-exponential distribution. We investigate the Bayesian D-optimal design for the unit exponential regression

Table 4: Nonparametric Bayesian D-optimal designs with uniform base distribution.

α	Prior distribution	Design	Two-point	
1	$U[1,10]$	x	0.08439	0.59184
		w	0.83337	0.16663
	$U[1,50]$	x	0.00241	0.28782
		w	0.99840	0.00160
	$U[1,100]$	x	0.00081	0.29072
		w	0.99685	0.00315
	$U[1,300]$	x	0.00012	0.30016
		w	0.99793	0.00207
5	$U[1,10]$	x	0.07819	0.11187
		w	0.70592	0.29408
	$U[1,50]$	x	0.00247	0.28584
		w	0.99884	0.00116
	$U[1,100]$	x	0.00066	0.29195
		w	0.99980	0.00020
	$U[1,300]$	x	0.00007	0.29538
		w	0.99930	0.00070
10	$U[1,10]$	x	0.07369	0.08374
		w	0.71886	0.28114
	$U[1,50]$	x	0.00248	0.28360
		w	0.99716	0.00284
	$U[1,100]$	x	0.00066	0.28980
		w	0.99958	0.00042
	$U[1,300]$	x	0.00007	0.29799
		w	0.99917	0.00083
50	$U[1,10]$	x	0.07456	0.39371
		w	0.82952	0.17048
	$U[1,50]$	x	0.00159	0.08261
		w	0.99993	0.00007
	$U[1,100]$	x	0.00062	0.19091
		w	0.99978	0.00022
	$U[1,300]$	x	0.00006	0.30198
		w	0.99966	0.00034

model (6) using a uniform prior distribution and examining various parameter values. By adopting a nonparametric Bayesian approach and utilizing the DP, we aim to address the challenges associated with selecting the prior distribution in Bayesian optimal design construction. This allows us to account for uncertainty and mitigate the impact of restrictive parametric assumptions providing more flexible and robust designs for nonlinear regression models.

In this investigation, our emphasis is on employing the uniform distribution as the base distribution in the DP. To provide a comprehensive understanding the influence of the concentration parameter α, we present the results in tables for four different values of $\alpha = 1, 5, 10, 50$. These tables provide valuable insights into the nonparametric Bayesian optimal designs and show the weights and support points. Through the analysis of results for different values of α, we can better understand the impact of this parameter on the design outcomes. This approach allows us to explore and evaluate the performance of the nonparametric Bayesian optimal designs across different levels of the concentration parameter α.

In the investigated range, the results reveal interesting findings. For all parameter values, the absence of three-point designs is notable. In fact, as uncertainty in the

base measure increases, another optimal point is obtained with a very small weight, forming a design where the two other point carry the highest weights. These designs categorized as two-point designs, given that the weight of the additional point becomes negligible.

As the uncertainty in the base measure and the concentration parameter in the DP increase, the support points in the two-point designs do not undergo significant changes. The weight of the smallest point increases rapidly, and it becomes the point with the highest weight. This weight tends to either increase or remain relatively stable with an increase in the concentration parameter.

It is crucial to highlight that the approach employed in this study holds applicability to other optimality criteria and diverse models featuring two or more parameters. For instance, extending nonparametric Bayesian optimal designs to encompass A- or E-optimality criteria for the discussed nonlinear model, in conjunction with a DP prior, presents avenues for future research. We anticipate presenting new findings in this domain in the near future.

References

Aarset, M.V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, **36**(1):106–108.

Aminnejad, M. and Jafari, H. (2017). Bayesian A and D-optimal designs for gamma regression model with inverse link function. *Communications in Statistics-Simulation and Computation*, **46**(10):8166–8189.

Atkinson, A., Donev, A. and Tobias, R. (2007). *Optimum Experimental Designs, with SAS*. Vol. 34, Oxford: OUP Oxford.

Bakouch, H.S., Hussain, T., Tošić, M., Stojanović, V.S. and Qarmalah, N. (2023). Unit exponential probability distribution: Characterization and applications in environmental and engineering data modeling. *Mathematics*, **11**(19):4207.

Burghaus, I. and Dette, H. (2014). Optimal designs for nonlinear regression models with respect to non-informative priors. *Journal of Statistical Planning and Inference*, **154**:12–25.

Bürkner, P.C., Schwabe, R. and Holling, H. (2019). Optimal designs for the generalized partial credit model. *British Journal of Mathematical and Statistical Psychology*, **72**(2):271–293.

Bondesson, L. (1982). On simulation from infinitely divisible distributions. *Advances in Applied Probability*, **14**(4):855–869.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, **24**(4): 586–602.

Chaloner, K.M. and Duncan, G.T. (1983). Assessment of a beta prior distribution: PM elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **32**(1-2):174–180.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, **21**(2):191–208.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, **10**(3):273-304.

Dette, H., Melas, V.B. and Wong, W.K. (2006). Locally D-optimal designs for exponential regression models. *Statistica Sinica*, **16**(3):789–803.

Dette, H. and Neugebauer, H.M. (1996). Bayesian optimal one point designs for one parameter nonlinear models. *Journal of Statistical Planning and Inference*, **52**(1):17–31.

Dette, H. and Neugebauer, H.M. (1997). Bayesian D-optimal designs for exponential regression models. *Journal of Statistical Planning and Inference*, **60**(2):331–349.

Fedorov, V.V. and Hackl, P. (2012). *Model-oriented Design of Experiments*. Vol. 125, Springer Science & Business Media.

Fedorov, V.V. and Leonov, S.L. (2013). *Optimal Design for Nonlinear Response Models*. CRC Press.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**(2):209–230.

Firth, D. and Hinde, J.P. (1997). On Bayesian D-optimum design criteria and the equivalence theorem in non-linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **59**(4):793–797.

Ford, I., Torsney, B. and Wu, C.J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **54**(2):569–583.

Ghitany, M.E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation*, **78**(4):493–506.

Goudarzi, M., Jafari, H. and Khazaei, S. (2021). Nonparametric Bayesian optimal designs for exponential regression model. *Communications in Statistics-Simulation and Computation*, **50**(6):1809–1819.

Graßhoff, U., Holling, H. and Schwabe, R. (2012). Optimal designs for the Rasch model. *Psychometrika*, **77**(4):710–723.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological)*, **21**(2):272–304.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, **30**(2):271–294.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, **30**(2):849–879.

Mazucheli, J., Menezes, A.F.B. and Chakraborty, S. (2019). On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *Journal of Applied Statistics*, **46**(4):700–714.

Mukhopadhyay, S. and Haines, L.M. (1995). Bayesian D-optimal designs for the exponential growth model. *Journal of Statistical Planning and Inference*, **44**(3):385–397.

Parsa Maram, P. and Jafari, H. (2016). Bayesian D-optimal design for logistic regression model with exponential distribution for random intercept. *Journal of Statistical Computation and Simulation*, **86**(10):1856–1868.

Pronzato, L. and Walter, É. (1985). Robust experiment design via stochastic approximation. *Mathematical Biosciences*, **75**(1):103–120.

Rodríguez-Torreblanca, C. and Rodríguez-Díaz, J.M. (2007). Locally D-and c-optimal designs for Poisson and negative binomial regression models. *Metrika*, **66**:161–172.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**(2):639–650.

Zarepour, M. and Al Labadi, L. (2012). On a rapid simulation of the Dirichlet process. *Statistics & Probability Letters*, **82**(5):916–924.