JIRSS (2026)
Vol. 25, No. 2, pp 213-229
DOI: 10.22034/jirss.2026.2022597.1049

Generalized unit-exponential distribution and its application to nonparametric Bayesian optimal design

Anita Abdollahi Nanvapisheh[1], **Habib Jafari**[1], **Soleiman Khazaei** [1]

[1]Department of Statistics, Razi University, Kermanshah, Iran
Received: 2024-02-10, Accepted: 2026-09-10, Published online: 2026-10-03

Abstract. This study introduces a novel two-parameter distribution, referred to as the *Generalized Unit-Exponential (G-UE) distribution*, specifically defined on the unit interval. The fundamental mathematical and statistical properties of the proposed distribution are thoroughly investigated. Parameter estimation is carried out using the method of maximum likelihood estimation (MLE). Furthermore, a special case of the proposed distribution obtained when $\gamma = 1$ is formulated as a nonlinear regression model. Nonlinear regression models have widespread applications across numerous scientific disciplines, including biostatistics, engineering, and the natural sciences. Achieving an accurate fit for such models is of critical importance, particularly when accounting for potential biases inherent in Bayesian optimal design frameworks. To address these challenges, this study develops a Bayesian optimal design methodology based on the Dirichlet process prior, which represents a fundamental tool in nonparametric Bayesian inference and provides several flexible and analytically tractable representations. The D-optimality criterion is employed in conjunction with the Dirichlet process as a functional prior mechanism to determine the nonparametric Bayesian optimal design.

Keywords. D-optimal design, Bayesian optimal design, Generalized Unit-Exponential model (G-UE), Dirichlet Process, Stick-Breaking Prior, Nonparametric Bayesian.

MSC: 62F15.

1 Introduction

In the field of experimental design, the concept of optimal design refers to a class of experimental strategies defined according to specific statistical optimality criteria. It is well established that a properly structured experiment can substantially enhance the precision and efficiency of statistical inference. In principle, prior to data collection for parameter estimation, it is advisable

Anita Abdollahi Nanvapisheh (anita.abdollahi@yahoo.com).
Corresponding Author: Habib Jafari (h.jafari@razi.ac.ir).
Soleiman Khazaei (s.khazaei@razi.ac.ir).

to determine the most appropriate experimental settings that maximize the information content of the data. This process, known as experimental design, becomes particularly crucial when prior information is limited and when data acquisition is costly in terms of time and financial resources.

Consequently, considerable research efforts have been devoted to addressing the problem of constructing optimal designs for nonlinear regression models. Experimental design plays a pivotal role across a wide range of scientific disciplines, including biomedicine and pharmacokinetics, where its application enables rigorous investigation and facilitates the extraction of meaningful and reliable scientific 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 et al., 1992; Rodríguez-Torreblanca and Rodríguez-Díaz, 2007), sequential optimal designs, minimax optimal designs, Bayesian optimal designs (Parsamaram and Jafari, 2016; Goudarzi et al., 2019; Grashoff et al., 2012; Kiefer, 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.

In the Bayesian framework, the initial step involves representing the available prior knowledge about the model parameters through a probability distribution, commonly referred to as the prior distribution. A Bayesian optimal design aims to maximize a specified optimality criterion averaged with respect to this prior distribution. However, it is important to recognize that the specification of the prior distribution may present substantial challenges and can potentially lead to misleading or biased results. The selection of the prior is inherently subjective, as it depends on the researcher's beliefs, expert knowledge, or historical information, and consequently may influence the final inference. Moreover, the Bayesian methodology does not provide a universally accepted or definitive procedure for determining the appropriate prior distribution. Numerous researchers such as Chaloner and Larntz (1989), Chaloner and Duncan (1983), Burghaus and Dette (2014), Chaloner and Verdinelli (1995), Pronzato and Walter (1985), Mukhopadhyay and Haines (1995), Dette and Neugebauer (1996), Dette and Neugebauer (1997), Fedorov and Hackl (2012), Fedorov and Leonov (2013), and Firth and Hinde (1997) have delved into investigating the impact of the prior distribution on determining design points in various types of optimal designs. Chapter 18 of Atkinson et al.'s book Atkinson et al. (2007) can be considered as a valuable resource in this regard. 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 paper introduces a novel two-parameter distribution, termed the *Generalized Unit-Exponential (G-UE) distribution*, specifically defined on the unit interval in Section 2. Subsequently, the proposed distribution is formulated as a nonlinear regression model, and its corresponding optimal designs are investigated. To this end, Section 3 begins with a review of optimal design theory for nonlinear regression models, followed by the derivation of the relevant theoretical results for the proposed model under a particular special case. Finally, Section 4 concludes the paper with a summary of the main findings and concluding remarks.

2 The Generalized Unit-Exponential distribution

If $Y \sim$ Exp(θ) with probability density function $f(y) = \theta \exp(-\theta y)$, 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 probability density function (pdf) of the resultant distribution (Unit Exponential distribution) are given by:

$$F(x \mid \theta) = 1 - \exp(\frac{-\theta x}{1-x});\ 0 \le x < 1,\ \theta > 0, \tag{2.1}$$

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

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

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

Now suppose that the random variable X has the cumulative distribution function of equation (2.1), in this case, the cumulative distribution function, the probability density function, and the hazard rate function of the Generalized Unit-Exponential (G-UE) distribution will be as follows:

$$F(x \mid \theta) = (1 - \exp(\frac{-\theta x}{1-x}))^{\gamma};\ 0 \le x < 1,\ \theta > 0, \gamma > 0,$$

$$f(x \mid \theta) = \gamma(\frac{\theta}{(1-x)^2} \exp(\frac{-\theta x}{1-x}))(1 - \exp(\frac{-\theta x}{1-x}))^{\gamma-1};\ 0 \le x < 1,\ \theta > 0, \gamma > 0,$$

$$h(x \mid \theta) = \frac{f(x \mid \theta)}{1 - F(x \mid \theta)} = \frac{\gamma(\frac{\theta}{(1-x)^2} \exp(\frac{-\theta x}{1-x}))(1 - \exp(\frac{-\theta x}{1-x}))^{\gamma-1}}{1 - (1 - \exp(\frac{-\theta x}{1-x}))^{\gamma}};\ 0 \le x < 1,\ \theta > 0, \gamma > 0.$$

In Figure 1, the probability density function and the hazard rate function 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 and U-shapes trend within the interval $0 \le x < 1$. In this Figure, both parameters were intentionally varied to illustrate their combined influence on the system behavior.

Moving forward, we will explore several key statistical properties of the G-UE distribution. Next, we shall discuss several main statistical properties of the G-UE distribution.

2.1 The moments and incomplete moments of the new distribution

Moments can be used to study the most important features and characteristics of a distribution. In this section, the complete and incomplete moments of the G-UE distribution are presented. First, we present an expansion for $f(x)$. Using the generalized binomial expansion:

$$(1-u)^{\alpha} = \sum_{j=0}^{\infty} \binom{\alpha}{j} (-1)^j u^j, \qquad |u| < 1.$$

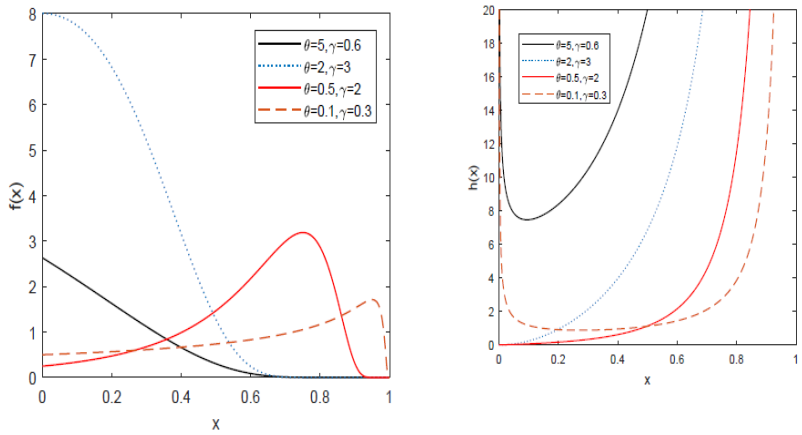


Figure 1: Plot of density function (left) and hrf (right)

Using the generalized binomial expansion with The probability density function (PDF) of the G-UE distribution can be expanded as

$$f(x \mid \theta) = \gamma \frac{\theta}{(1-x)^2} \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \exp\left(\frac{-\theta(j+1)x}{1-x}\right).$$

Using this series expansion, the cumulative distribution function (CDF) is obtained by integrating term by term:

$$\begin{aligned} F(x \mid \theta) &= \int_0^x f(t \mid \theta)\, dt \\ &= \int_0^x \gamma \frac{\theta}{(1-t)^2} \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \exp\left(\frac{-\theta(j+1)t}{1-t}\right) dt \\ &= \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \gamma\theta \int_0^x \frac{1}{(1-t)^2} \exp\left(\frac{-\theta(j+1)t}{1-t}\right) dt. \end{aligned}$$

By the substitution $u = \frac{t}{1-t} \implies t = \frac{u}{1+u}, dt = \frac{du}{(1+u)^2}$, the integral simplifies to

$$F(x \mid \theta) = \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \left[1 - \exp\Big(-\theta(j+1)\frac{x}{1-x}\Big)\right].$$

Thus, the CDF of the G-UE distribution has the series representation

$$F(x \mid \theta) = \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \left[1 - \exp\Big(-\theta(j+1)\frac{x}{1-x}\Big)\right], \quad 0 < x < 1.$$

So, the moment generating function of the model using (8) is given by:

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \int_0^1 e^{tx} f(x \mid \theta)\, dx.$$

By substituting the series representation of $f(x \mid \theta)$, we have

$$M_X(t) = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \int_0^1 \frac{e^{tx}}{(1-x)^2} \exp\left(\frac{-\theta(j+1)x}{1-x}\right) dx.$$

Using the change of variable

$$u = \frac{x}{1-x} \quad \Rightarrow \quad x = \frac{u}{1+u}, \quad dx = \frac{du}{(1+u)^2}, \quad \frac{1}{(1-x)^2} = (1+u)^2,$$

the integral becomes

$$M_X(t) = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \int_0^{\infty} \exp\left(\frac{tu}{1+u} - \theta(j+1)u\right) du.$$

Hence, the moment generating function of the G-UE distribution is

$$M_X(t) = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \int_0^{\infty} \exp\left(\frac{tu}{1+u} - \theta(j+1)u\right) du.$$

The r-th moment of the G-UE distribution, from (8) is given by:

$$\mu'_r = \mathbb{E}[X^r] = \int_0^1 x^r f(x \mid \theta)\, dx.$$

Substituting the series expansion of $f(x \mid \theta)$, we have

$$\mu'_r = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \int_0^1 \frac{x^r}{(1-x)^2} \exp\left(\frac{-\theta(j+1)x}{1-x}\right) dx.$$

Using the change of variable $u = \frac{x}{1-x} \Rightarrow x = \frac{u}{1+u}, dx = \frac{du}{(1+u)^2}, \frac{1}{(1-x)^2} = (1+u)^2$, the integral becomes

$$\mu'_r = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \int_0^{\infty} \left(\frac{u}{1+u}\right)^r \exp\big(-\theta(j+1)u\big) du.$$

which can be easily calculated using Maple software. Now, The skewness (Sk) and kurtosis (K) of the distribution are given by:

$$\mu'_r = \mathbb{E}[X^r] = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \int_0^{\infty} \left(\frac{u}{1+u}\right)^r \exp(-\theta(j+1)u) du.$$

Skewness

The skewness of X is defined as

$$\text{Skewness} = \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3} = \frac{\mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3}{(\mu'_2 - (\mu'_1)^2)^{3/2}},$$

where $\mu = \mu'_1 = \mathbb{E}[X]$ and $\sigma^2 = \mu'_2 - (\mu'_1)^2$.

Kurtosis

The kurtosis of X is defined as

$$\text{Kurtosis} = \frac{\mathbb{E}[(X-\mu)^4]}{\sigma^4} = \frac{\mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2\mu_2' - 3(\mu_1')^4}{(\mu_2' - (\mu_1')^2)^2}.$$

The excess kurtosis is

$$\text{Excess Kurtosis} = \text{Kurtosis} - 3.$$

Let $X \sim$ G-UE(γ, θ). The r-th *incomplete moment* up to x is defined by

$$m_r(x) = \int_0^x t^r f(t \mid \theta)\, dt.$$

Substituting the series expansion of $f(t \mid \theta)$, we get

$$m_r(x) = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \int_0^x \frac{t^r}{(1-t)^2} \exp\left(\frac{-\theta(j+1)t}{1-t}\right) dt.$$

Using the change of variable $u = \frac{t}{1-t} \Rightarrow t = \frac{u}{1+u}, dx = \frac{du}{(1+u)^2}, \frac{1}{(1-t)^2} = (1+u)^2$, the integral becomes

$$m_r(x) = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \int_0^{\frac{x}{1-x}} \left(\frac{u}{1+u}\right)^r \exp\left(-\theta(j+1)u\right) du, \quad 0 < x < 1.$$

2.2 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). Let $X \sim$ G-UE(γ, θ) with mean $\mu = \mathbb{E}[X] = \mu_1'$. The mean deviation around the mean is defined as

$$\text{MD} = \mathbb{E}[|X-\mu|] = \int_0^1 |x-\mu| f(x \mid \theta) dx.$$

Dividing the integral at μ, we have

$$\text{MD} = \int_0^{\mu} (\mu - x) f(x \mid \theta) dx + \int_{\mu}^1 (x-\mu) f(x \mid \theta) dx.$$

Substituting the series representation of $f(x \mid \theta)$, we obtain

$$MD = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \left[\int_0^{\mu} (\mu - x) \frac{1}{(1-x)^2} \exp\left(\frac{-\theta(j+1)x}{1-x}\right) dx + \int_{\mu}^1 (x-\mu) \frac{1}{(1-x)^2} \exp\left(\frac{-\theta(j+1)x}{1-x}\right) dx \right].$$

Using the change of variable $u = \frac{x}{1-x} \Rightarrow x = \frac{u}{1+u}, dx = \frac{du}{(1+u)^2}, \frac{1}{(1-x)^2} = (1+u)^2$, the mean deviation becomes

$$MD = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma-1}{j} (-1)^j \left[\int_0^{\frac{\mu}{1-\mu}} \left(\mu - \frac{u}{1+u}\right) e^{-\theta(j+1)u} du + \int_{\frac{\mu}{1-\mu}}^{\infty} \left(\frac{u}{1+u} - \mu\right) e^{-\theta(j+1)u} du \right].$$

Now let $X \sim$ G-UE(γ, θ) with median $\tilde{x}$ satisfying $F(\tilde{x} \mid \theta) = 0.5$. The mean deviation about the median is defined as

$$\mathrm{MD}_{\mathrm{median}} = \mathbb{E}[|X - \tilde{x}|] = \int_0^1 |x - \tilde{x}| f(x \mid \theta) dx.$$

Dividing the integral at $\tilde{x}$, we have

$$\mathrm{MD}_{\mathrm{median}} = \int_0^{\tilde{x}} (\tilde{x} - x) f(x \mid \theta) dx + \int_{\tilde{x}}^1 (x - \tilde{x}) f(x \mid \theta) dx.$$

Substituting the series expansion of $f(x \mid \theta)$, we obtain

$$\mathrm{MD}_{\mathrm{median}} = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma - 1}{j} (-1)^j \left[\int_0^{\tilde{x}} (\tilde{x} - x) \frac{1}{(1-x)^2} \exp\Big(\frac{-\theta(j+1)x}{1-x}\Big) dx + \int_{\tilde{x}}^1 (x - \tilde{x}) \frac{1}{(1-x)^2} \exp\Big(\frac{-\theta(j+1)x}{1-x}\Big) dx \right].$$

Using the change of variable $u = \frac{x}{1-x} \Rightarrow x = \frac{u}{1+u}, dx = \frac{du}{(1+u)^2}, \frac{1}{(1-x)^2} = (1+u)^2$, the mean deviation about the median becomes

$$\mathrm{MD}_{\mathrm{median}} = \gamma\theta \sum_{j=0}^{\infty} \binom{\gamma - 1}{j} (-1)^j \left[\int_0^{\frac{\tilde{x}}{1-\tilde{x}}} \left(\tilde{x} - \frac{u}{1+u} \right) e^{-\theta(j+1)u} du + \int_{\frac{\tilde{x}}{1-\tilde{x}}}^{\infty} \left(\frac{u}{1+u} - \tilde{x} \right) e^{-\theta(j+1)u} du \right].$$

2.3 Maximum Likelihood Estimation

In this section, we will explore the estimation of the parameter θ for the G-UE distribution using both maximum likelihood methodology and the method of moments. Suppose $X_1, X_2, \ldots, X_n$ form a random sample from the G-UE distribution with the cdf given in (2.1). The log-likelihood function of θ can then be expressed as:

$$\ell(\theta, \gamma \mid \boldsymbol{x}) \propto n \log \gamma + \sum_{i=1}^{n} \log\Big(\frac{\theta}{(1-x_i)^2} \exp\Big(\frac{-\theta x_i}{1-x_i}\Big)\Big) + (\gamma - 1) \sum_{i=1}^{n} \log\Big(1 - \exp\Big(\frac{-\theta x_i}{1-x_i}\Big)\Big), \qquad (2.2)$$

The ML estimates of θ and γ are obtained by maximizing the log-likelihood function with respect to the parameters. To do this, we take the derivatives of the log-likelihood function with respect to θ and γ and then equate the results with zero. Therefore, we have:

$$\frac{\partial \ell(\theta, \gamma \mid \boldsymbol{x})}{\partial \theta} = \sum_{i=1}^{n} \left(\frac{1}{\theta} - \frac{x_i}{1-x_i} \right) + (\gamma - 1) \sum_{i=1}^{n} \left[\frac{x_i}{1-x_i} \cdot \frac{\exp\left(\frac{-\theta x_i}{1-x_i}\right)}{1 - \exp\left(\frac{-\theta x_i}{1-x_i}\right)} \right] = 0$$

$$\frac{\partial \ell(\theta, \gamma \mid \boldsymbol{x})}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^{n} \log\left(1 - \exp\left(\frac{-\theta x_i}{1-x_i}\right)\right) = 0,$$

Then:

$$\hat{\gamma}(\theta) = \frac{n}{-\sum_{i=1}^{n} \log\left(1 - \exp\left(\frac{-\theta x_i}{1-x_i}\right)\right)}$$

and:

$$\frac{n}{\hat{\theta}} - \sum_{i=1}^{n} \frac{x_i}{1-x_i} + \left(\hat{\gamma}(\hat{\theta}) - 1\right) \sum_{i=1}^{n} \frac{\frac{x_i}{1-x_i} \exp\left(\frac{-\hat{\theta} x_i}{1-x_i}\right)}{1 - \exp\left(\frac{-\hat{\theta} x_i}{1-x_i}\right)} = 0$$

We may use numerical methods to find the solutions of the above equations. Next, we obtain the observed information matrix that is useful for finding approximate confidence intervals for the parameters.This matrix is defined as follows:

$$I_n(\Omega) = -\begin{bmatrix} I_{\theta\theta} & I_{\theta\gamma} \\ I_{\theta\gamma} & I_{\gamma\gamma} \end{bmatrix},$$

where $\Omega = (\theta, \gamma)^T$ is the vector of parameters and:

$$I_{\gamma\gamma} = \frac{\partial^2 \ell(\theta, \gamma)}{\partial \gamma^2},$$
$$I_{\theta\gamma} = \frac{\partial^2 \ell(\theta, \gamma)}{\partial \theta \, \partial \gamma},$$
$$I_{\theta\theta} = \frac{\partial^2 \ell(\theta, \gamma)}{\partial \theta^2}.$$

Let $J(\Omega)^{-1}$ be the inverse of the expected Fisher information matrix and $\widehat{\Omega}$ be the ML estimator of Ω. If the sample size n is large enough, then under the regularity conditions, $\sqrt{n}(\widehat{\Omega} - \Omega)$ follows a two-variate normal distribution asymptotically with the mean $(0, 0)^T$ and the variance-covariance matrix $J(\Omega)^{-1}$. We may replace $J(\Omega)^{-1}$ with $\left[\frac{1}{n} I_n(\Omega)\right]^{-1}$ as the asymptotic behavior remains valid. The unknown parameters that may appear in $\left[\frac{1}{n} I_n(\Omega)\right]^{-1}$ can be replaced by their corresponding ML estimates. We may obtain approximate (asymptotic) confidence intervals for the parameters and perform test hypotheses using the above explained normal approximation.

2.4 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 Generalized Unit Exponential, Unit-Rayleigh, Unit-Weibull, and Unit-BXII density, we compute LogLik, the Akaike Information Criterion (*AIC*), and the Bayesian Information Criterion (*BIC*). The results are summarized in Table 3.

The optimal model corresponds to lower *LogLik*, AIC and BIC values. The values of these measures suggest that the G-UE distribution is a robust contender against other competitive distributions and, moreover, stands out as the best fit among them. Figure 3 illustrates the

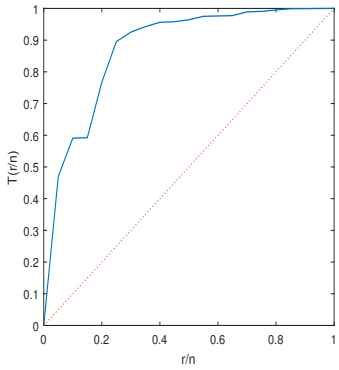


Figure 2: TTT plot for the data Set

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 G-UE distribution is a suitable model for fitting the considered dataset. Figure 3 visually confirms the superior fit of the Best Distribution by demonstrating its PDF curve most accurately traces the empirical histogram's shape, especially near the peak and in the tails, indicative of matching skewness/kurtosis. This visual assessment is statistically substantiated by achieving the minimum value across quantitative metrics, such as the AIC/BIC criterion. Consequently, this distribution provides the most robust model for the underlying data generating process compared to alternatives.

Table 1: Parameter estimates, log-likelihood values and goodness-of-fit measures for the data set.

Model	Parameters	LogLik	AIC	BIC	KS stat	p-value
GUE	$\hat{\alpha} = 2.9725,\ \hat{\theta} = 0.5329$	15.6025	-27.2051	-25.2136	0.187	0.434
Exponential	$\hat{\lambda} = 1.3864$	-13.4657	28.9314	29.9271	0.419	0.001
Gamma	$\hat{k} = 17.7949,\ \widehat{scale} = 0.0405$	7.3246	-10.6492	-8.6578	0.291	0.054
Weibull	$\hat{k} = 7.2987,\ \widehat{scale} = 0.7748$	12.2673	-20.5347	-18.5432	0.222	0.239

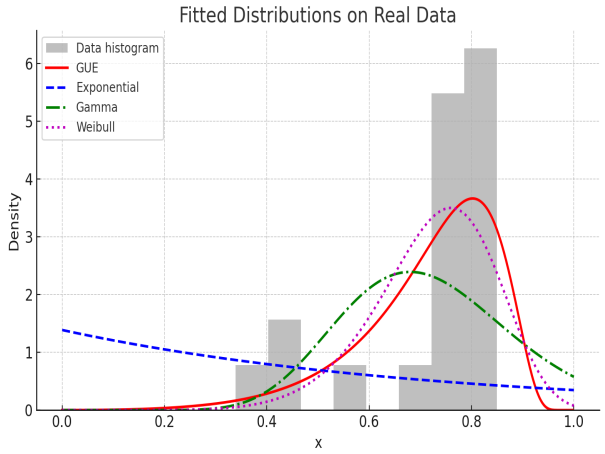


Figure 3: the histogram of the dataset alongside the fitted density function and plots of the empirical and estimated pdf for these fitted distributions.

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 $x \in \chi \subseteq \mathrm{R}$ 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 et al. (2019) and Abdollahi et al. (2023), and Abdollahi et al. (2024):

$$I(\xi, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \eta(x, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}^T} \eta(x, \boldsymbol{\theta}). \tag{3.1}$$

And the information function of ξ can be expressed as follows Atkinson et al. (2007):

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

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:

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

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:

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

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

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:

$$\Psi_{\Pi}(\xi) = E(\psi(\xi; \boldsymbol{\theta})) = \int_{\Theta} \psi(\xi; \boldsymbol{\theta}) d\Pi(\boldsymbol{\theta}) = \int_{\Theta} -\log(\det(M(\xi, \boldsymbol{\theta}))) d\Pi(\boldsymbol{\theta}), \tag{3.2}$$

where Π represents the prior distribution for $\boldsymbol{\theta}$ and the Bayesian D-optimal design is attained by minimizing Equation (3.2).

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 Dirichlet processes. 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 Ferguson (1973) that defined a random probability measure which is a Dirichlet process on ($\boldsymbol{\Theta}$, B($\boldsymbol{\Theta}$)), as:

$$P(.) = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}(.), \tag{3.3}$$

where θ_i ($i > 1$) is a sequence of *i.i.d.* random variables with common distribution Q and δ_{θ_i} represents a probability measure that is degenerate at θ that δ_{θ_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 Dirichlet process. Bondesson (1982), Sethuraman (1994), and Zarepour and Al Labadi (2012) are among those who have contributed to this area. A method of producing samples from the Dirichlet process is to use the Polya urn process that in the upcoming section, we will discuss about it. Then the nonparametric Bayesian D-optimal design for the UE model is discussed.

3.2 Polya Urn Scheme

Polya Urn Scheme was used by Blackwell and McQueen (1973) to demonstrate the existence of the Dirichlet process. The method of producing a sample of the Dirichlet process is to use a Polya Urn Scheme Ghosh and Ramamoorthi (2003). Consider a Polya urn with $a(\chi)$ balls of which $a(i)$ are of color i ; $i = 1, 2, ..., k$.[For the moment assume that $a(i)$ are whole numbers or 0]. Draw balls at random from the urn, replacing each ball drawn by two balls of the same color. Let $X_i = j$ if the i th ball is of color j. Then:

$$P(X_1 = j) = \frac{a(j)}{a(\chi)},$$

$$P(X_2 = j \mid X_1) = \frac{a(j) + \delta_{X_1}(j)}{a(\chi) + 1},$$

and in general

$$P(X_{n+1} = j \mid X_1, X_2, ..., X_n) = \frac{a(j) + \sum_{1}^{n} \delta_{X_i}(j)}{a(\chi) + n},$$

That n is the number of extracted balls and $\delta_{X_i}(j)$ is equal to one if $X_i = j$, otherwise it is equal to zero.

3.2.1 Nonparametric Bayesian D-optimal design for G-UE nonlinear model in a special case (γ=1)

Now, let's consider nonparametric Bayesian D-optimal design for G-UE nonlinear model in a special case (γ=1) as following:

$$E(y|x) = \eta(x, \boldsymbol{\theta}) = \frac{\theta}{(1-x)^2} \exp(\frac{-\theta x}{1-x}), 0 \leq x < 1, \theta > 0. \tag{3.4}$$

Therefore, the Bayesian D-optimality criterion, denoted as $\Psi_\Pi(\xi)$ can be expressed according to our model as follows Goudarzi et al. (2019) and Abdollahi et al. (2025):

$$\Psi_\Pi(\xi) = E(\psi(\xi; \boldsymbol{\theta})) = \int_\Theta -\log(\sum_{j=1}^{\ell} w_j[\exp(\frac{-\theta x_j}{1-x_j})(\frac{1}{(1-x_j)^2} + \frac{\theta x_j}{(1-x_j)^3})]^2) d\Pi(\theta) \tag{3.5}$$

where Π is the prior distribution for θ. The Bayesian D-optimal design is attained by minimizing Equation (3.5). 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:

$$\Psi_\Pi(\xi) = \sum_{i=1}^{\infty} p_i(-\log(\sum_{j=1}^{\ell} w_j[\exp(\frac{-\theta_i x_j}{1-x_j})(\frac{1}{(1-x_j)^2} + \frac{\theta_i x_j}{(1-x_j)^3})]^2)). \tag{3.6}$$

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. Chaloner and Duncan (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 [0, 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 (3.6) represents a stochastic function of the DP. According to Ferguson's definition of the DP, the calculation of Equation (3.6) 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.

Now, in this section we consider Polya Urn Scheme as the base measure in DP. We get the results by using a nonlinear optimization programing with R package Rsolnp. To better understanding of the effect of the α parameter, we tabulate the results for four different values of α=1, 5, 10, 50, in Tables 2–5. We also fixed ϵ=10^{-10}. Without loss of generality, we consider

Table 2: Nonparametric Bayesian D-optimal design when α=1.

Prior distribution	*Design*	*Two − point*		*Three − point*		
$U[0,10]$	x	0.17838	0.68105	0.11980	0.37940	0.79569
	w	0.68675	0.31325	0.94547	0.04676	0.00777
$U[0,50]$	x	0.00476	0.22862	0.00285	0.19403	0.32482
	w	0.94189	0.05811	0.97973	0.00016	0.02011
$U[0,100]$	x	0.00216	0.20343	0.00106	0.19451	0.30591
	w	0.96776	0.03224	0.98745	0.00005	0.01250
$U[0,300]$	x	0.00049	0.19904	0.00025	0.19661	0.29978
	w	0.98160	0.01840	0.99625	0.00025	0.00350

Table 3: Nonparametric Bayesian D-optimal design when α=5.

Prior distribution	*Design*	*Two − point*		*Three − point*		
$U[0,10]$	x	0.14566	0.70363	0.06769	0.43661	0.70570
	w	0.60655	0.39345	0.88170	0.03461	0.08369
$U[0,50]$	x	0.00546	0.24275	0.00240	0.18749	0.31624
	w	0.92609	0.07391	0.97558	0.00080	0.02362
$U[0,100]$	x	0.00421	0.21048	0.00093	0.19097	0.30643
	w	0.96830	0.03170	0.98873	0.00030	0.01097
$U[0,300]$	x	0.00026	0.20151	0.00018	0.19700	0.29981
	w	0.98843	0.01157	0.99752	0.00035	0.00213

a bounded design space χ=[0, 1]. Tables 2–5 represent 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 many cases, researchers are interested in comparing designs obtained by various criteria. The concept of design efficiency is utilized to compare two designs. In this paper, we used the relative efficiency as presented by Chaloner and Larntz (1989). The results of investigating the efficiency of a two-point optimal design compared to the corresponding three-point optimal design for B= 10, 50 considering the uniform base distribution and α=10 are presented in Table 6. A relative efficiency value closer to one indicates that the ξ optimal design will provide as much information as the ξ^* optimal design.

Table 4: Nonparametric Bayesian D-optimal design when α=10.

Prior distribution	*Design*	*Two − point*		*Three − point*		
$U[0,10]$	x	0.13553	0.68324	0.09643	0.34264	0.75747
	w	0.61186	0.38814	0.73415	0.10585	0.16000
$U[0,50]$	x	0.00496	0.24662	0.00262	0.18789	0.32683
	w	0.91923	0.08077	0.97641	0.00144	0.02215
$U[0,100]$	x	0.00093	0.20362	0.00095	0.19130	0.30914
	w	0.98026	0.01974	0.98871	0.00100	0.01029
$U[0,300]$	x	0.00015	0.20004	0.00014	0.19442	0.30217
	w	0.98528	0.01472	0.99311	0.00044	0.00645

Table 5: Nonparametric Bayesian D-optimal designs with uniforml base distribution and concentration parameter when α=50. First row: support points; second row: weights.

Prior distribution	*Design*	*Two points*		*Three points*		
$U[0,10]$	x	0.11600	0.65172	0.11297	0.40796	0.73798
	w	0.62599	0.37401	0.68415	0.14066	0.17519
$U[0,50]$	x	0.00287	0.24340	0.00259	0.18591	0.34163
	w	0.93935	0.06065	0.95704	0.00091	0.04205
$U[0,100]$	x	0.00092	0.21214	0.00084	0.19041	0.30341
	w	0.96278	0.03722	0.99082	0.00025	0.00892
$U[0,300]$	x	0.00012	0.19973	0.00013	0.19580	0.30657
	w	0.99214	0.00786	0.99418	0.00024	0.00558

Table 6: Efficiency of the two- and three-point nonparametric Bayesian D-optimal design compared to each other for $p = 2$ and $\alpha = 10$.

Prior distribution	Design	Two-point		Three-point		efficiency
$U[0,10]$	x	0.13553	0.68324	0.09643	0.34264	0.75747
	w	0.61186	0.38814	0.73415	0.10585	0.16000
$U[0,50]$	x	0.00496	0.24662	0.00262	0.18789	0.32683
	w	0.91923	0.08077	0.97641	0.00144	0.02215

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 model (Equation (3.4)) 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 α=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. The results show that in the class of two- and three-point designs, the largest weight corresponds to the smallest point. Moreover, as the uncertainty in the base measure and the concentration parameter in the Dirichlet process increase, the support points in the two- and three-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 Dirichlet process prior, presents avenues for future research. We anticipate presenting new findings in this domain in the near future.

References

Aarset MV. How to identify a bathtub hazard rate. IEEE Transactions on Reliability. 1987;36(1):106–108. https://doi.org/10.1109/TR.1987.5222310.

Abdollahi A, Jafari H, Khazaei S. Nonparametric Bayesian optimal designs for unit exponential nonlinear model. Journal of Statistical Modelling: Theory and Applications. 2023;.

Abdollahi A, Jafari H, Khazaei S. Nonparametric Bayesian Optimal Designs for Unit Expo-

nential Regression Model with Respect to Prior Processes (with the Truncated Normal as the Base Measure). Statistics in Transition. 2024;.

Abdollahi A, Jafari H, Khazaei S. Nonparametric Bayesian optimal designs for unit exponential regression model with respect to prior processes (with Polya Urn scheme as the base measure). International Journal of Nonlinear Analysis and Applications. 2025;.

Aminnejad M, Jafari H. Bayesian A- and D-optimal designs for gamma regression model with inverse link function. Communications in Statistics - Simulation and Computation. 2017;46(10):8166–8189. https://doi.org/10.1080/03610918.2016.1271888.

Atkinson AC, Donev AN, Tobias RD. Optimum Experimental Design, with SAS. Oxford, UK: Oxford University Press; 2007.

Blackwell D, McQueen J. Discreteness of Ferguson selections. Annals of Statistics. 1973;1(2):356–358. https://doi.org/10.1214/aos/1176342373.

Bondesson L. On simulation from infinitely divisible distributions. Advances in Applied Probability. 1982;14(4):855–869. https://doi.org/10.2307/1426771.

Burghaus I, Dette H. Optimal designs for nonlinear regression models with respect to non-informative priors. Journal of Statistical Planning and Inference. 2014;154:12–25. https://doi.org/10.1016/j.jspi.2014.05.009.

Bürkner PC, Schwabe R, Holling H. Optimal designs for the generalized partial credit model. British Journal of Mathematical and Statistical Psychology. 2019;72(3):405–429. https://doi.org/10.1111/bmsp.12148.

Chaloner K, Duncan GT. Assessment of a beta prior distribution: PM elicitation. The Statistician. 1983;32(1/2):174–180. https://doi.org/10.2307/2987609.

Chaloner K, Larntz K. Optimal Bayesian design applied to logistic regression experiments. Journal of Statistical Planning and Inference. 1989;21(2):191–208. https://doi.org/10.1016/0378-3758(89)90004-9.

Chaloner K, Verdinelli I. Bayesian experimental design: a review. Statistical Science. 1995;10(3):273–304. https://doi.org/10.1214/ss/1177009939.

Chernoff H. Locally optimal designs for estimating parameters. The Annals of Mathematical Statistics. 1953;24(4):586–602. https://doi.org/10.1214/aoms/1177728915.

Dette H, Melas VB, Wong WK. Locally D-optimal designs for exponential regression. Statistica Sinica. 2006;16(3):789–803.

Dette H, Neugebauer HM. Bayesian optimal one point designs for one parameter nonlinear models. Journal of Statistical Planning and Inference. 1996;52(1):17–31. https://doi.org/10.1016/0378-3758(95)00104-2.

Dette H, Neugebauer HM. Bayesian D-optimal designs for exponential regression models. Journal of Statistical Planning and Inference. 1997;60(2):331–349. https://doi.org/10.1016/S0378-3758(96)00138-3.

Fedorov V, Hackl P. Model-Oriented Design of Experiments, vol. 125. Springer Science and Business Media; 2012.

Fedorov V, Leonov SL. Optimal Design for Nonlinear Response Models. CRC Press; 2013.

Ferguson TS. A Bayesian analysis of some nonparametric problems. The Annals of Statistics. 1973;1(2):209–230. https://doi.org/10.1214/aos/1176342360.

Firth D, Hinde J. On Bayesian D-optimum Design Criteria and the Equivalence Theorem in Non-linear Models. Journal of the Royal Statistical Society B. 1997;59(4):793–797. https://doi.org/10.1111/1467-9868.00096.

Ford I, Torsney B, Wu CFJ. 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. 1992;54(2):569–583.

Ghitany ME, Atieh B, Nadarajah S. Lindley distribution and its application. Mathematics and Computers in Simulation. 2008;78(4):493–506. https://doi.org/10.1016/j.matcom.2007.06.007.

Ghosh JK, Ramamoorthi RV. Bayesian Nonparametrics. New York: Springer-Verlag; 2003.

Goudarzi M, Jafari H, Khazaei S. Nonparametric Bayesian optimal designs for exponential regression model. Communications in Statistics - Simulation and Computation. 2019;50(11):3323–3339. https://doi.org/10.1080/03610918.2019.1593454.

Grashoff U, Holling H, Schwabe R. Optimal designs for the Rasch model. Psychometrika. 2012;77(4):710–723. https://doi.org/10.1007/s11336-012-9276-2.

Kiefer J. Optimum experimental designs. Journal of the Royal Statistical Society B. 1959;21(2):272–319.

Kiefer J. General equivalence theory for optimum designs (approximate theory). The Annals of Statistics. 1974;2(5):849–879. https://doi.org/10.1214/aos/1176342810.

Mazucheli J, Menezes AFB, Chakraborty S. On the one parameter unit-Lindley distribution and its associated regression model for proportion data. Journal of Applied Statistics. 2019;46(4):700–714. https://doi.org/10.1080/02664763.2018.1511774.

Mukhopadhyay S, Haines LM. Bayesian D-optimal designs for the exponential growth model. Journal of Statistical Planning and Inference. 1995;44(3):385–397. https://doi.org/10.1016/0378-3758(94)00073-5.

Parsamaram P, Jafari H. Bayesian D-optimal Design for the logistic regression model with exponential distribution for random intercept. Journal of Statistical Computation and Simulation. 2016;86(9):1856–1868. https://doi.org/10.1080/00949655.2015.1087525.

Pronzato L, Walter E. Robust experiment design via stochastic approximation. Mathematical Biosciences. 1985;75(1):103–120. https://doi.org/10.1016/0025-5564(85)90068-9.

Rodríguez-Torreblanca C, Rodríguez-Díaz JM. Locally D-optimal designs for the heteroscedastic linear model. Computational Statistics and Data Analysis. 2007;51(10):4899–4912. https://doi.org/10.1016/j.csda.2006.09.003.

Sethuraman J. A Constructive Definition of Dirichlet Priors. Statistica Sinica. 1994;4(2):639–650.

Zarepour M, Al Labadi L. On a rapid simulation of the Dirichlet process. Statistics and Probability Letters. 2012;82(5):916–924. https://doi.org/10.1016/j.spl.2012.01.020.