Journal of Statistical Planning and Inference 215 (2021) 193–207

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

D-optimal designs for estimation of parameters in a simplex dispersion model with proportional data

Hsiang-Ling Hsu [a,*], Mong-Na Lo Huang [b]

[a] Institute of Statistics, National University of Kaohsiung, Kaohsiung, Taiwan, ROC
[b] Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan, ROC

ARTICLE INFO

Article history:
Received 21 March 2020
Received in revised form 13 February 2021
Accepted 5 March 2021
Available online 22 March 2021

Keywords:
Link function
Proportional data
Statistical inference
Weight function
Weighted polynomial regression

ABSTRACT

In this work, optimal design problems for estimation of unknown parameters for a flexible class of non-normal distributions useful for describing various data types are considered. A particular model, designated the simplex dispersion model, can be applied to model proportional (or compositional) outcomes confined within the (0, 1) interval. The main interest here is to determine the optimal experimental settings to be able to estimate the unknown model parameters more accurately and efficiently. Locally D-optimal designs for accurate estimation of parameters in the simplex dispersion model are characterized through the corresponding equivalence theorem and under certain cases with some given prior information, optimal design results are presented for illustration. Examples including a water purification experiment and a dose study are used to demonstrate the efficiencies of the corresponding D-optimal designs.

1. Introduction

Continuous proportional outcomes where responses are confined within the interval (0, 1) are observed and collected from many practical studies. This type of data, also named as percentage or ratio data, is acquired quite often in public health or environmental sciences. In environmental study or drug analysis, we may be concerned with the percentage of reduction on observed response values such as the water purification rate or cure rate etc. The main purpose of this paper is to investigate related optimal design problems for estimation of parameters of models suitable for proportional data. Wu et al. (2005) and Latif and Zafar Yab (2015) had considered design problems for modeling proportional response data based on beta regression with one variable x restricted to [0, 1].

For analyzing binary response data, the generalized linear model (GLM) with binomial or logistic distributions are often used to model the corresponding response probability by distribution functions taking values within (0, 1). Here instead, a special type of distribution in the class of dispersion models (DM), defined by Jørgensen (1997) and introduced later in Section 2 will be used, which is a very appealing family of distributions to model the proportional type of data considered here. The distribution in this class, denoted as the "simplex dispersion model (SD model, $S^{-}(\mu, \sigma^2)$)", has support within the (0, 1) interval, rendering it be a possible choice for characterizing continuous proportional data. Some works on longitudinal studies have employed this type of model; see Song and Tan (2000) and Qiu et al. (2008). Song and Tan (2000) demonstrated that the variance–covariance matrix of the estimator of the unknown parameters in the SD model is formed by a function containing the model parameters. In what follows, we consider the locally optimal design problems and omit the word "locally" for simplicity.

* Corresponding author.
E-mail address: hsuhl@nuk.edu.tw (H.-L. Hsu).

https://doi.org/10.1016/j.jspi.2021.03.002
0378-3758/

In the development of optimal design theory, weighted polynomial regression models with variance functions depending on the explanatory variable have played an important role. In the literature on optimal design theory about estimation of linear models, the differential equation methodology for weighted polynomial regression models was utilized initially in Karlin and Studden (1966), Huang et al. (1995), Chang and Lin (1997), Imhof et al. (1988), and Dette et al. (1999) thereafter. For the polynomial regression model without an intercept, Huang et al. (1995) treated the design problem as that for a special type of weighted polynomial model, and derived appropriate differential equations and solved corresponding eigenvalue problems for obtaining the optimal solution. See Sitter and Torsney (1995a,b), and Torsney and Musrati (1993) among others for the discussions on D-optimal designs under the weighted models. Using geometric and other arguments, Musrati (1992) reviewed and augmented the methods of optimal designs for non-linear problems with a single variable. Dette and Trampisch (2010) unified different types of D-optimal design problems in weighted, univariate polynomial regression models for a broad class of efficiency functions. Here the construction of D-optimal designs for simplex dispersion models can also be viewed as that for weighted polynomial regression model with a special type of variance which has not been studied before. The corresponding weight function for the SD model has an interesting feature which can be expressed as the sum of two functions where one is proportional to the inverse of the other.

The rest of the paper is organized as follows. Preliminaries about dispersion models and optimal designs for nonlinear models are presented in Section 2. Section 3 centers around the characteristics of the weight function of the corresponding information matrix and the D-optimal designs for continuous proportional data with linear link functions, respectively, in the simplex dispersion model. Examples including a water purification experiment and a dose study are illustrated in Section 4. Finally, discussion and conclusions are presented.

2. Preliminaries: Simplex dispersion models and *D*-optimality

Jørgensen (1997) introduced a class of dispersion models to extend the known exponential family models. This particular class of distribution functions extends the Euclidean distance $(y-\mu)^2$ in the normal density to a general discrepancy function $\underline{d}(y;\mu)$. Each of such distributions is determined by the discrepancy function, and the resulting distribution is fully parameterized by the location (or mean) parameter μ and dispersion parameter σ. A formal definition for dispersion models is given as follows; also see Song (2007).

Definition 1. A (reproductive) dispersion model $\mathrm{DM}(\mu,\sigma^2)$ with location parameter μ and dispersion parameter σ^2 is a family of distributions whose probability density functions take the following form:

$$p(y;\mu,\sigma^2)=a(y;\sigma^2)\exp\{-\frac{1}{2\sigma^2}\underline{d}(y;\mu)\},\ \ y\in\mathcal{C},$$

where $\mu\in\Omega,\ \sigma^2>0$, and $a(y;\sigma^2)\geq 0$ is a suitable normalizing term that is independent of μ. Usually $\Omega\subseteq\mathcal{C}\subseteq\mathcal{R}$ and $\mathcal{C}$ is the support of the density.

The bivariate function $\underline{d}(\cdot;\cdot)$ is called the unit deviance defined on $(y,\mu)\in\mathcal{C}\times\Omega$. The unit variance function $V:\Omega\to(0,\infty)$ is

$$V(\mu)=\frac{2}{\frac{\partial^2}{\partial\mu^2}\underline{d}(y;\mu)|_{y=\mu}}\ .$$

Note that for the SD model, $a(y;\sigma^2)=[2\pi\sigma^2\{y(1-y)\}^3]^{-1/2}$, $\underline{d}(y;\mu)=\frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}$, $V(\mu)=\mu^3(1-\mu)^3$ and $\mathcal{C},\ \Omega=(0,1)$ which are suitable for continuous proportional data. As shown in Figure 2.1 of Song (2007), the SD model can be adopted for modeling with more flexible characteristics on the mean μ and dispersion σ^2 parameters. Jørgensen (1997) formulated the variance of y in the SD model (also see Song and Tan (2000)) as

$$\mathrm{Var}(y)=\mu(1-\mu)-\frac{1}{\sqrt{2}\sigma}\exp\left\{\frac{1}{\sigma^2\mu^2(1-\mu)^2}\right\}\Gamma\left\{\frac{1}{2},\frac{1}{2\sigma^2\mu^2(1-\mu)^2}\right\},$$

where $\Gamma(a,b)=\int_b^\infty x^{a-1}e^{-x}dx$ is the incomplete gamma function. Although the SD model is extended from the normal density function, the variance of y in SD model is not equal to σ^2. For this together with other properties and examples of dispersion models, and for more explanation and details, see Jørgensen (1997) and Song (2007).

It is important to understand firstly the structure of the information matrix for the parameter estimates to find the optimal designs. In addition to the dispersion models, the maximum likelihood estimates for the unknown model parameters under the specific link function is presented in Song (2007) as well. Here, we briefly summarize the formulations and properties needed for characterizing the D-optimal designs in the following.

Assume that a data set $(y_i,x_i),\ i=1,\ldots,n$, where the responses y_i's are independent from $S^-(\mu_x,\sigma^2)$ distribution where μ_x depends on the corresponding covariate variable x, i.e.,

$$\ln(\mu_x/(1-\mu_x))=f^{\top}(x)\boldsymbol{\alpha} \tag{1}$$

For $S^{-}(\mu_x, \sigma^2)$ with the parameter vector $\boldsymbol{\theta}^{\mathsf{T}} = (\boldsymbol{\alpha}^{\mathsf{T}}, \sigma^2)$, let $\boldsymbol{y}, \tilde{\boldsymbol{x}}$ denote $(y_1, \ldots, y_n)^{\mathsf{T}}, (x_1, \ldots, x_n)^{\mathsf{T}}$, respectively. Then the Fisher information matrix for $\boldsymbol{\theta}$ based on the corresponding log-likelihood function is a block diagonal matrix, i.e.,

$$I(\boldsymbol{\theta}, \tilde{\boldsymbol{x}}) = \begin{pmatrix} I_{\alpha,\alpha}(\boldsymbol{\theta}, \tilde{\boldsymbol{x}}) & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & I_{\sigma^2,\sigma^2}(\boldsymbol{\theta}, \tilde{\boldsymbol{x}}) \end{pmatrix}. \tag{2}$$

Since the first order derivative of the deviance score is $\delta(y_i; \mu_i) = -\frac{1}{2}\frac{\partial d(y_i;\mu_i)}{\partial \mu_i}$ and $E(\frac{\partial d(y_i;\mu_i)}{\partial \mu_i}) = 0$ (see Chapter 2 in Song (2007)), we have that

$$I_{\alpha,\sigma^2}(\boldsymbol{\theta}, \tilde{\boldsymbol{x}}) = -E\left(\sum_{i=1}^{n} \frac{1}{(\sigma_i^2)^2} \frac{f(\boldsymbol{x}_i)}{\dot{g}(\mu_i)} \delta(y_i; \mu_i)\right) = \mathbf{0} = I_{\sigma^2,\alpha}(\boldsymbol{\theta}, \tilde{\boldsymbol{x}}),$$

where $\dot{g}(\mu_i)$ is the first order derivative of link function g w.r.t. μ_i. Thus, (2) merely involves two elements

$$I_{\alpha,\alpha}(\boldsymbol{\theta}, \tilde{\boldsymbol{x}}) = \boldsymbol{X}_f^{\mathsf{T}} U^{-1} \boldsymbol{X}_f, \quad \text{and} \quad I_{\sigma^2,\sigma^2}(\boldsymbol{\theta}, \tilde{\boldsymbol{x}}) = \frac{n}{2\sigma^4} \tag{3}$$

where $\boldsymbol{X}_f$ is a matrix with the ith row being the $f^{\mathsf{T}}(x_i)$, U is a diagonal matrix with ith diagonal element ν_i/σ_i^{-2}, and $\nu_i = (\dot{g}(\mu_i))^2/E(-\dot{\delta}(y_i; \mu_i))$, where $\dot{\delta}(y_i; \mu_i)$ is the first order derivative of deviance score $\delta(y_i; \mu_i)$.

Throughout this study we consider approximate design ξ of the following form

$$\xi = \begin{Bmatrix} x_1 & \cdots & x_r \\ w_1 & \cdots & w_r \end{Bmatrix}, \quad 0 < w_i \le 1, \sum_{i=1}^{r} w_i = 1,$$

which is a probability measure on a compact design space $\mathcal{X}$ with a finite number of support points. Consequently, the normalized information matrix for $\boldsymbol{\theta}$ is $M(\boldsymbol{\theta}, \xi) = \int_{\mathcal{X}} I(\boldsymbol{\theta}, \tilde{\boldsymbol{x}})\xi(dx)$, dependent on the unknown parameter vector $\boldsymbol{\theta}$. This makes investigation of the optimal design problems more complicated. Search of optimal designs requires an initial guess and prior information of the unknown parameters in the model. D-optimal designs minimize the determinant of the inverse of the information matrix $M(\boldsymbol{\theta}, \xi)$ among the set of all feasible approximate designs on design space. The well known equivalence theorem in optimal design theory corresponding to the D-optimality described by White (1973) and Whittle (1973), and the optimality of candidate designs will be verified through the corresponding theorem. In practice, both location and dispersion parameters may be unknown but some prior information may be available. As aforementioned, the information matrix for $\boldsymbol{\alpha}$ and σ^2 is a diagonal matrix and the D-optimality criterion seeking a design ξ to maximize $|M(\boldsymbol{\theta}, \xi)|$ is equivalent to maximizing $|M(\boldsymbol{\alpha}, \xi)| = |\int_{\mathcal{X}} I_{\alpha,\alpha}(\boldsymbol{\theta}, \tilde{\boldsymbol{x}})\xi(dx)|$ with respect to ξ, whether σ is known or not. We assume that σ is a known constant parameter for elaborating the investigation.

3. *D*-optimal designs for linear link function with two parameters

Now we concentrate on the D-optimality criterion and investigate the case with linear logit link function in one variable x, namely $f(x) = \boldsymbol{x} = (1\ x)^{\mathsf{T}}$ and

$$\ln(\mu_x/(1-\mu_x)) = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{\alpha} = \alpha_0 + \alpha_1 x = z. \tag{4}$$

According to the canonical transformation approach as in Ford et al. (1992), Sitter and Torsney (1995b) and Yang et al. (2011), it is useful to transform the design point x to the point z as follows

$$\boldsymbol{x} = \begin{pmatrix} 1 & 0 \\ -\alpha_0/\alpha_1 & 1/\alpha_1 \end{pmatrix} \begin{pmatrix} 1 \\ z \end{pmatrix} = \boldsymbol{A}_{\alpha}\boldsymbol{z}.$$

Usually the range of z can be constructed to be within specific lower and upper bounds and can be transformed further such that $-b \le z \le b$, which yields the design space $\mathcal{X}_b = [-b, b]$ for some $b \ge 0$. That is for $L \le z' \le U$, let $\boldsymbol{z} = (1\ z)^{\mathsf{T}} = B\left(1\ z'\right)^{\mathsf{T}}$ such that $z \in \mathcal{X}_b$ where B is a nonsingular matrix of order 2 with elements (L, U, b). Now, denote $\xi_{\boldsymbol{z}}$ and $\xi_{\boldsymbol{x}}$ as designs based on $\boldsymbol{z_i}$ and $\boldsymbol{x}_i$, $i = 1, \ldots, r$, respectively. Following the set up above, the information matrix based on (2) and (4) with the known constant dispersion parameter can be expressed as

$$M(\boldsymbol{\alpha}, \xi_{\boldsymbol{x}}) = \boldsymbol{A}_{\boldsymbol{\alpha}} M(\boldsymbol{\alpha}, \xi_{\boldsymbol{z}}) \boldsymbol{A}_{\boldsymbol{\alpha}}^{\mathsf{T}}, \tag{5}$$

where $M(\boldsymbol{\alpha}, \xi_{\boldsymbol{x}}) = \sum_{i=1}^{r} w_i I_{\alpha,\alpha}(\boldsymbol{\theta}, x_i)$ and $M(\boldsymbol{\alpha}, \xi_{\boldsymbol{z}}) = \sum_{i=1}^{r} w_i \lambda(z_i) f(z_i) f^{\mathsf{T}}(z_i)$, $f(z_i) = (1\ z_i)^{\mathsf{T}}$, where $\lambda(z_i)$ is named as the weight function at z_i in the later discussion.

3.1. Properties of the weight function

The weight function with $c_\sigma = 1/(3\sigma^2)$ and $\mu_i = 1/(1+e^{-z_i})$ for $S^{-}(\mu, \sigma^2)$ has definition

$$\lambda(z_i) = \frac{\nu_i^{-1}}{\sigma^2} = 3\tilde{\lambda}(z_i) = 3\Big(\lambda_1(z_i) + \lambda_2(z_i)\Big) = 3\left(\mu_i(1-\mu_i) + \frac{c_\sigma}{\mu_i(1-\mu_i)}\right), \tag{6}$$

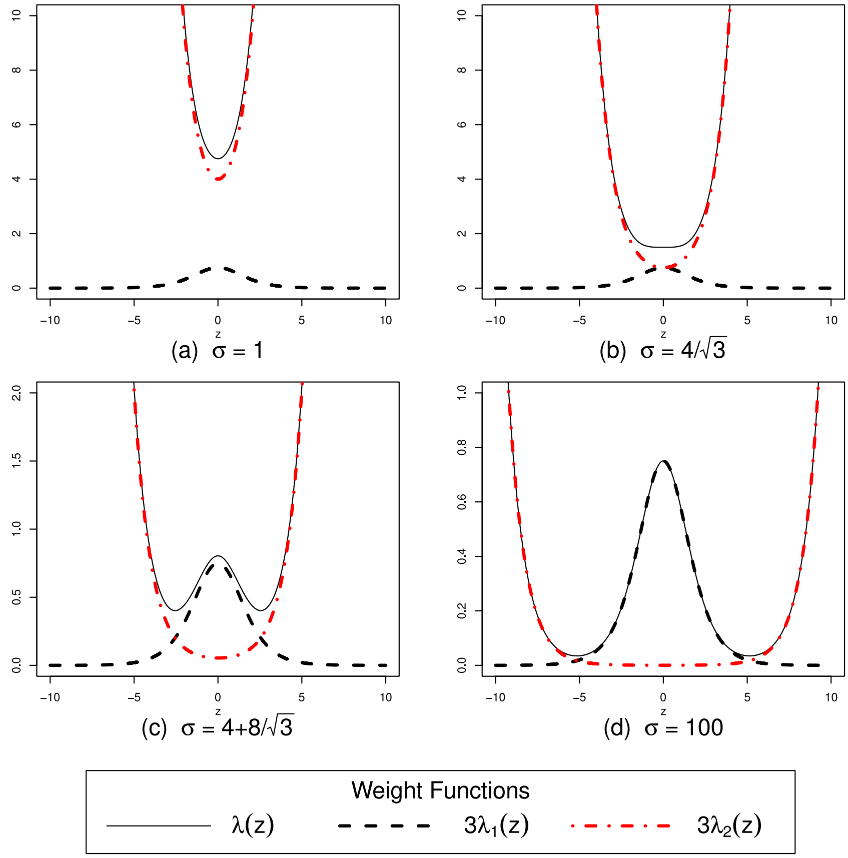

Fig. 1. The graphs of functions $\lambda(z)$, $3\lambda_1(z)$ and $3\lambda_2(z)$ for $\sigma = 1, 4/\sqrt{3}, 4+8/\sqrt{3}, 100$.

which is different from those in the other commonly used GLMs, e.g., logit or probit binary response models. Note that (6) is also given in Song (2007, p. 42–44). More explicitly, (6) is the sum of two functions related to the mean at the design points as well as the common variance. This kind of structure is quite interesting and deserves special attention. The patterns of the weight function and its subfunctions are illustrated in Fig. 1 and characterized in Appendix A to facilitate our study.

Fig. 1 shows that the behavior of $\lambda(z)$, which is affected mainly by $\lambda_2(z)$ when σ is small, and when σ gets larger as well as $\mathcal{X}_b$ is restricted within a specific compact set, e.g., $\sigma = 100$ and $b \leq 4$, $\lambda(z)$ turns to be dominated by $\lambda_1(z)$.

Before demonstrating more properties about the weight function, some relevant notations should be defined here. Let $\sigma_1^* = \arg\min_{\sigma>0}\{\lambda_1^2(z) \leq c_\sigma,\ \forall z\} = 4/\sqrt{3}(\approx 2.309)$, and $z_0 = \arg\left\{z : \lambda_1^2(z) - c_\sigma = 0,\ z > 0 \text{ and } \sigma > \sigma_1^*\right\}$, namely,

$$z_0 = \log\left[-1 + \frac{1+\sqrt{1-4\sqrt{c_\sigma}}}{2\sqrt{c_\sigma}}\right] = \log\left[-1 + \frac{\sqrt{3}\sigma + \sqrt{\sigma(3\sigma - 4\sqrt{3})}}{2}\right]. \tag{7}$$

Note that Jørgensen (1997) had proved that the simplex distribution has a uni-mode if $\sigma < 4/\sqrt{3} = \sigma_1^*$; otherwise, it yields multi-modes. Examining the pattern of $\lambda(z)$ for $z \geq 0$ as illustrated in Lemma A.4, $z = 0$ is a minimum when $\sigma \leq \sigma_1^*$ and is the only local maximum when $\sigma > \sigma_1^*$.

To this end, let the function

$$\tau_y(x) = \lambda(x)/\lambda(y) \tag{8}$$

be the ratio of $\lambda(x)$ and $\lambda(y)$ for $x \neq y$ and $\dot{\lambda}(z)$ be the first derivative of $\lambda(z)$. Then, we can summarize some useful properties about $\tau_y(x)$ with respect to σ, b and will make use of these results for characterizing possible candidates for

D-optimal designs. Furthermore, let

$$b^* = \arg\min_{z>0}\{z : \lambda(z) \geq \lambda(0), \forall \sigma > \sigma_1^*\} = \log\left[\frac{1}{8}\left(-8 + 3\sigma^2 + \sqrt{-48\sigma^2 + 9\sigma^4}\right)\right];$$
$$\sigma_2^* = \arg\max_{\sigma>0}\{\sigma : \lambda(z) \geq \lambda(0)/2,\ \forall z \in \mathbb{R}\} = 4 + 8/\sqrt{3}(\approx 8.619),$$

we have the following results.

Proposition 1. *For function $\lambda(z)$ defined in (6), we have the properties if*

(i) $\{\sigma \leq \sigma_2^*\}$, *then* $\tau_0(z) \geq 1/2,\ \forall z \in \mathcal{X}_b$; *moreover, when* $\{\sigma \leq \sigma_1^*\}$ *or* $\{\sigma_1^* < \sigma \leq \sigma_2^*, b \geq b^*\}$, *then* $\tau_b(z) \leq 1,\ \forall z \in \mathcal{X}_b$;

(ii) $\{\sigma > \sigma_2^*\}$,

- *(a) when* $b \geq b^*$, *then* $\tau_b(z) \leq 1,\ \forall z \in \mathcal{X}_b$;
- *(b) there are two solutions* z_1, z_2 *for* $\tau_0(z) = 1/2,\ z > 0$, *where* $0 < z_1 < z_2$ *are given explicitly in* (B.4) and (B.5);
- *(c) moreover with* $b \geq z_2$, *then* $\tau_b(z) \leq b/z,\ \forall z \in (0, b]$; *on the other hand if* $b \leq z_1$, *there exists at most one* a^*, $0 < a^* \leq b$, *such that*

$$\begin{aligned} &\dot{\lambda}(a^*)a^* + \lambda(a^*) = 0, \\ &\tau_z(a^*) \geq z/a^*,\ \forall z \in (0, b]. \end{aligned} \tag{9}$$

Proposition 1 and the behavior of $\lambda(z)$ would facilitate characterizing the patterns of candidate designs, where values of σ_1^*, σ_2^* determine the shape of the weight function $\lambda(z)$, exhibiting the relative influences of $\lambda_1(z)$ and $\lambda_2(z)$ to the optimal designs. The ratio functions $\tau_0(z)$ and $\tau_b(z)$ as well as the values of b^*, z_1 and z_2 obtained above help classify ranges of b, the end point of the design interval, with respect to the number of support points of the corresponding optimal designs.

3.2. Characterizing the D-optimal designs

From , we know that to maximizing $|M(\boldsymbol{\alpha}, \xi_{\boldsymbol{x}})|$ with respect to the design $\xi_{\boldsymbol{x}}$ is equivalent maximizing $|M(\boldsymbol{\alpha}, \xi_{\boldsymbol{z}})|$ with respect to $\xi_{\boldsymbol{z}}$ whether σ is known or not. Subsequently, we will focus on the problem of finding designs which maximize $|M(\boldsymbol{\alpha}, \xi_{\boldsymbol{z}})|$. Denote $\xi_{\boldsymbol{z}}$ by ξ for brevity. It is easy to verify that $[\mu_i(1-\mu_i)]$ and $\lambda(z_i)$ in (6) are even functions. The invariance theorems in Kiefer (1959, 1961) and Giovagnoli et al. (1987) are very useful in restricting investigations to the class of symmetric designs, as there always exists a D-optimal design which is symmetric.

The terms (σ, b), hence, need to be taken into account. They create major difficulties in characterizing optimal designs in a SD model. Restricted within the class of symmetric designs, denoted as Ξ_b, we provide a procedure on how to search for a candidate design first and later verify its optimality by the equivalence theorem.

Let $\bar{\xi}$ be a symmetric design with $2\bar{r}$ supports in Ξ_b as

$$\bar{\xi} = \begin{Bmatrix} -z_{\bar{r}} & \cdots & -z_1 & z_1 & \cdots & z_{\bar{r}} \\ w_{\bar{r}}/2 & \cdots & w_1/2 & w_1/2 & \cdots & w_{\bar{r}}/2 \end{Bmatrix},\ \text{where } z_i \in [0, b],\ w_i > 0,\ \sum_{i=1}^{\bar{r}} w_i = 1.$$

Based on the symmetric property of $\lambda(z)$, the information matrix is simplified as

$$M(\boldsymbol{\alpha}, \bar{\xi}) = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}$$

through the two elements M_1 and M_2. Moreover the dispersion function of $\bar{\xi}$ is of the form

$$d(z, \bar{\xi}) = \lambda(z)[\kappa_1(\bar{\xi}) + \kappa_2(\bar{\xi})z^2] \tag{10}$$

where $\lambda(z)$ is as defined in (6), and $\kappa_1(\bar{\xi}),\ \kappa_2(\bar{\xi}) > 0$ are the diagonal elements of the inverse of the information matrix of $\bar{\xi}$. In order to find the support points of the optimal designs, we need to find the points achieving the maximum value of the dispersion function within $\mathcal{X}_b$. So these interior support points of the optimal design will have zero derivative of (10).

Lemma 1. *In Ξ_b, there exists a symmetric D-optimal design with at most four support points, and there are at most two interior support points that are symmetric with respect to zero ($\pm a$, $0 \leq a < b$), satisfying*

$$\frac{\dot{\lambda}(z)}{\lambda(z)} = -\frac{2z}{\kappa(\bar{\xi}) + z^2}, \quad \kappa(\bar{\xi}) = \frac{\kappa_1(\bar{\xi})}{\kappa_2(\bar{\xi})}. \tag{11}$$

In Lemma 1, it is interesting that in (11), on the left hand side of the equation, the behavior of the function depends only on the weight function, while on the right hand side, it depends on the design. On the other hand, we can maximize $|M(\boldsymbol{\alpha}, \bar{\xi})| = M_1M_2$ by selecting $\bar{r}$, z_i and w_i for $i = 1, \ldots, \bar{r}$. Note that (M_1, M_2) generates a 2-dimensional convex hull spanned by $\{(\lambda(z), y) : 0 \leq y \leq \lambda(z)z^2,\ 0 \leq z \leq \infty\}$. Therefore from Lemma 1 or with a similar argument based on the Caratheodory's theorem in Silvey (1980) and Sitter and Wu (1993, Section 3), it follows that we can consider designs with $\bar{r}$ equal to 1 or 2, yielding that $\bar{\xi}$ contains at most four support points. Moreover, as the dispersion function in (10) can have at most one local maximum within $[0, b]$, this also yields that, if a symmetric optimal design has three support points, it can only be supported at zero and the two end points (i.e., $\{-b, 0, b\}$).

In the following, we divide our investigations into two steps: (i) examine the patterns of the corresponding determinant of the information matrix under three candidate symmetric designs with supports at two end points and no more than four supports totally; (ii) present conditions on the model parameters to identify the D-optimal designs. To this end, we consider symmetric designs with at most four support points including the two end points. Let $S = \{(a, \gamma) \in [0, b) \times [0, 1]\} \setminus (0, 1)$ and a four point design $\bar{\xi}_{a,\gamma}$ be

$$\bar{\xi}_{a,\gamma} = \begin{Bmatrix} -b & -a & a & b \\ (1-\gamma)/2 & \gamma/2 & \gamma/2 & (1-\gamma)/2 \end{Bmatrix}, \ (a, \gamma) \in S.$$

Note that $\bar{\xi}_{a,\gamma}$ would be reduced to a two point design when γ equals 0 or 1, and reduced to a three point design when $a = 0$, $\gamma \in (0, 1)$. For convenience, denote (i) $\bar{\xi}_{a,0}$ as $\bar{\xi}^2_b$, (ii) $\bar{\xi}_{a,1}$ as $\bar{\xi}^2_a$, (iii) $\bar{\xi}_{0,\gamma}$ as $\bar{\xi}^3_\gamma$ respectively.

Step (i): Properties of the D-optimality function

Now for any design $\bar{\xi}_{a,\gamma}$, $(a, \gamma) \in S$, the determinant of $M(\boldsymbol{\alpha}, \bar{\xi}_{a,\gamma})$ is proportional to a function of (a, γ) denoted by

$$\begin{aligned} g(a, \gamma) =& (\gamma\lambda(a) + (1-\gamma)\lambda(b))(\gamma\lambda(a)a^2 + (1-\gamma)\lambda(b)b^2) \\ =& (\lambda(b) - \lambda(a)) \left(\lambda(b)b^2 - \lambda(a)a^2\right) \gamma^2 + \left(\lambda(a)\lambda(b)(a^2 + b^2) - 2\lambda^2(b)b^2\right) \gamma \\ &+ \lambda^2(b)b^2. \end{aligned} \tag{12}$$

Let γ^*_a denote the optimal γ for given a, where the maximum value of $g(a, \gamma)$, $0 \le \gamma \le 1$ occurs, that is

$$\gamma^*_a = \arg \max_{0 \le \gamma \le 1} \{g(a, \gamma)\}. \tag{13}$$

Locating the value of γ^*_a, as presented in Lemma 2 through the ratio functions $\tau_a(b)$ defined in (8), will be beneficial for identifying potential candidate designs. In what follows, Lemma 2 characterizes the conditions for the better designs, e.g., if b satisfies Lemma 2(i) for all $a \in (0, b)$, $\bar{\xi}_2$ is D-optimal.

Lemma 2. *For $0 \le \gamma \le 1$, $\tau_a(b) = \lambda(b)/\lambda(a)$ and a given a, $0 \le a < b$, γ^*_a defined in (13) has the following properties*

(i) when $0 < a < b$,

(a) $\gamma^*_a = 0$ if $(a^2 + b^2)/(2b^2) \le \tau_a(b)$;
(b) $\gamma^*_a = 1$ if $\tau_a(b) \le 2a^2/(a^2 + b^2)$;
(c) $0 < \gamma^*_a < 1$ if $2a^2/(a^2 + b^2) < \tau_a(b) < (a^2 + b^2)/(2b^2)$.

(ii) when $a = 0$,

$$\gamma^*_0 = \begin{cases} \frac{2\tau_0(b)-1}{2(\tau_0(b)-1)} & \text{if } \tau_0(b) \le 1/2. \\ 0 & \text{otherwise}; \end{cases} \tag{14}$$

Step (ii): Determination of the D-optimal designs with model parameters

Making use of Proposition 1 and Lemma 2, we can acquire and verify an admissible design among $\bar{\xi}^2_b$, $\bar{\xi}^2_a$ and $\bar{\xi}^3_\gamma$, presented in the following theorems where the proofs are given in Appendix C.

Theorem 1 (*Better Designs*)**.** *For the SD model with dispersion parameter σ and design region $\mathcal{X}_b$, $b \in \mathbb{R}^+$, and for z_1, z_2, a^* as given in Proposition 1,*

(i) $\{0 < \sigma \le \sigma^*_2\}$ or $\{\sigma > \sigma^*_2, b \ge z_2\}$, $|M(\boldsymbol{\alpha}, \bar{\xi}^2_b)|$ is the maximum, i.e., the design $\bar{\xi}^2_b$ is better than the other three designs;
(ii) $\{\sigma > \sigma^*_2, b \le z_1\}$, $|M(\boldsymbol{\alpha}, \bar{\xi}^2_{a^*})|$ is the maximum, i.e., $\bar{\xi}^2_{a^*}$ is better than other designs if there exists $a^* < b$, where a^* is as satisfying (9). Otherwise, $|M(\boldsymbol{\alpha}, \bar{\xi}^2_b)|$ is the maximum, which means that $\bar{\xi}^2_b$ is better than other designs.
(iii) $\{\sigma > \sigma^*_2, b \in (z_1, z_2)\}$ and γ^*_0 as given in (14), then $|M(\boldsymbol{\alpha}, \bar{\xi}^3_{\gamma^*_0})| > |M(\boldsymbol{\alpha}, \bar{\xi}^2_b)|$, i.e., the design $\bar{\xi}^3_{\gamma^*_0}$ is better than $\bar{\xi}^2_b$.

Note that based on conditions listed in Proposition 1, five categories of pairs of (b, σ) are characterized and together with results from Lemma 2, Theorem 1 can be obtained after some straightforward computation. According to Theorem 1, the types of admissible designs under different categories of pairs of (b, σ) are presented in Table 1 and exhibited in Fig. 2, where region labels follow those in Theorem 1.

From Theorem 1(iii), for which the pair (b, σ) falls into region (iii), as marked in Fig. 2, we need to compare further whether $\bar{\xi}^2_{a^*}$, $\bar{\xi}^3_{\gamma^*_0}$ or one of the four point symmetric designs in $\{\bar{\xi}_{a,\gamma}\}$ would be optimal. Further studies yield the last characterization of optimal designs, which completes investigations and are presented in Theorem 2.

Theorem 2 (*Optimal Designs*)**.** *For the SD model with dispersion parameter σ and design region $\mathcal{X}_b$, $b \in \mathbb{R}^+$, if*

Table 1
The corresponding categories of the candidate designs under different pairs of (b, σ).

Categories on (b,σ)		Candidate designs
$\sigma \le \sigma_2^*$	$\sigma < \sigma_1^*$	$\bar{\xi}_b^2$
	$\sigma > \sigma_1^*$	$\bar{\xi}_b^2$
$\sigma > \sigma_2^*$	$b \le z_1$	$\{\bar{\xi}_a^2, \bar{\xi}_b^2\}$
	$b \in (z_1, z_2)$	$\{\bar{\xi}_a^2, \bar{\xi}_\gamma^3, \bar{\xi}_{a,\gamma}\}$
	$b \ge z_2$	$\bar{\xi}_b^2$

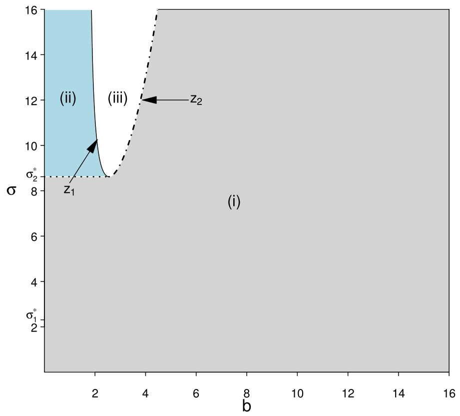

Fig. 2. The corresponding regions of the candidate optimal designs under different pairs of (b, σ).

(i) $\{0 < \sigma \le \sigma_2^*\}$ *or* $\{\sigma > \sigma_2^*, b \ge z_2\}$, $\bar{\xi}_b^2$ *is D-optimal;*

(ii) $\{\sigma > \sigma_2^*, b \le z_1\}$, $\bar{\xi}_{a^*}^2$ *is D-optimal if additionally there exists* $a^* < b$, *with* a^* *satisfying* (9). *Otherwise,* $\bar{\xi}_b^2$ *is D-optimal;*

(iii) $\{\sigma > \sigma_2^*, b \in (z_1, z_2)\}$, *where* z_1, z_2 *are as given in* (B.4) *and* (B.5), *design* $\bar{\xi}_{a^*_{\gamma^*},\gamma^*}$ *is D-optimal, if additionally there exists a solution pair* $(a^*_{\gamma^*}, \gamma^*) \in (0, b) \times (0, 1]$ *with*

$$\gamma = \frac{2b^2 - \tau_b(a)(a^2 + b^2)}{2(1 - \tau_b(a))(b^2 - \tau_b(a)a^2)}, \tag{15}$$

and with γ *given in* (15),

$$a^*_\gamma = arg\left\{a : \frac{\partial g(\gamma, a)}{\partial a} = 0, 0 < a < b\right\},$$

i.e., $a^*_{\gamma^*}$ *satisfies* $\tau_b(a)[\gamma(\dot{\lambda}(a)a + \lambda(a)) + 1] - \dot{\lambda}(a)\lambda(a) = 0$; *otherwise, there exists a* γ_0^* *as defined in* (14), *such that design* $\bar{\xi}^3_{\gamma_0^*}$ *is D-optimal.*

Numerical approximate D-optimal designs with various (b, σ) for the SD model with linear logit link function are presented in Table 2. From the results obtained, there are some interesting features of the optimal designs. From Theorem 2(i), when σ is small, that is $\sigma < \sigma_2^* = 8.619$, $\bar{\xi}_b$ is to be D-optimal design. When σ is relatively large but b is not large enough to fall into regions (ii) and (iii) shown in Fig. 2, we need to compare the four types of designs $\bar{\xi}_b^2$, $\bar{\xi}_a^2$, $\bar{\xi}_\gamma^3$ and $\bar{\xi}_{a,\gamma}$ to confirm the D-optimal design. In Table 2, for $\sigma = 15$, then $b = 1.500$ or 4.340 falls into the case of Theorem 2(ii) and (i), respectively; on the other hand for $b = 1.900$ or 2.300 or 3.200 we have that $b \in (z_1, z_2)$ as in Theorem 2(iii). As mentioned in Section 3.1, when σ is large enough, $\lambda_1(z)$ plays a dominating role, and it reveals that the limiting design, when σ goes to infinity, the D-optimal design is approximating the D-optimal design in the logistic regression model with binary responses (see Minkin (1987)) on some compact design spaces. In Table 2 with large σ, e.g. $\sigma = 1000$, on certain specific compact design space, the D-optimal design, equally supported at ± 1.5434, is nearly the D-optimal design in the logistic regression model with binary responses. However from Theorem 1, the region of the design space also plays an important role and the supports of the D-optimal design will eventually be the two boundary points $\{-b, b\}$, when b gets larger.

The D-efficiency of a candidate design ξ is defined as

$$D_{\text{eff}}(\bar{\xi}) = \left(\frac{|M(\boldsymbol{\alpha}, \bar{\xi})|}{|M(\boldsymbol{\alpha}, \bar{\xi}^*)|}\right)^{1/s}, \tag{16}$$

Table 2
The approximate D-optimal designs for $\sigma = 10, 15, 1000$ and different b.

σ	a^*	z_1	z_0	z_2	b^*	b	ξ^*	$D_{\text{eff}}(\bar{\xi}^U_{4,b})$
10	–	2.096	2.725	3.303	4.290	2.000	$\begin{Bmatrix} -2.00000 & 2.00000 \\ 0.50000 & 0.50000 \end{Bmatrix}$	0.906
						2.100	$\begin{Bmatrix} -2.10000 & -0.37726 & 0.37726 & 2.10000 \\ 0.49918 & 0.00082 & 0.00082 & 0.49918 \end{Bmatrix}$	0.917
						2.350	$\begin{Bmatrix} -2.35000 & 0.00000 & 2.35000 \\ 0.46390 & 0.07219 & 0.46390 \end{Bmatrix}$	0.934
						3.400	$\begin{Bmatrix} -3.40000 & 3.40000 \\ 0.50000 & 0.50000 \end{Bmatrix}$	0.847
15	1.718	1.877	3.176	4.326	5.116	1.500	$\begin{Bmatrix} -1.50000 & 1.50000 \\ 0.50000 & 0.50000 \end{Bmatrix}$	0.857
						1.900	$\begin{Bmatrix} -1.71826 & 1.71826 \\ 0.50000 & 0.50000 \end{Bmatrix}$	0.920
						2.300	$\begin{Bmatrix} -2.30000 & -1.69370 & 1.69370 & 2.30000 \\ 0.02927 & 0.47073 & 0.47073 & 0.02927 \end{Bmatrix}$	0.961
						3.200	$\begin{Bmatrix} -3.20000 & 0.00000 & 3.20000 \\ 0.35759 & 0.28483 & 0.35759 \end{Bmatrix}$	0.983
						4.340	$\begin{Bmatrix} -4.34000 & 4.34000 \\ 0.50000 & 0.50000 \end{Bmatrix}$	0.804
1000	1.543	1.763	7.456	12.834	13.528	1.000	$\begin{Bmatrix} -1.00000 & 1.00000 \\ 0.50000 & 0.50000 \end{Bmatrix}$	0.798
						1.560	$\begin{Bmatrix} -1.54343 & 1.54343 \\ 0.50000 & 0.50000 \end{Bmatrix}$	0.881
						10.000	$\begin{Bmatrix} -10.00000 & -1.31925 & 1.31925 & 10.00000 \\ 0.10036 & 0.39964 & 0.39964 & 0.10036 \end{Bmatrix}$	0.470
						11.500	$\begin{Bmatrix} -11.50000 & 0.00000 & 11.50000 \\ 0.28791 & 0.42419 & 0.28791 \end{Bmatrix}$	0.447
						13.000	$\begin{Bmatrix} -13.00000 & 13.00000 \\ 0.50000 & 0.50000 \end{Bmatrix}$	0.524

where s is the number of unknown parameters. For example, let $\bar{\xi}^U_{4,b}$ be the uniform design with four supports, equally spaced in $\mathcal{X}_b$, with equal weights. The D-efficiencies of $\bar{\xi}^U_{4,b}$ with different (b, σ) are also illustrated in Table 2. Most of these efficiencies are greater than 0.79, but the worst case is lower than 0.45, when b and σ are simultaneously large.

4. Example

Now, we present two examples, based on water treatment data and mortality data, to illustrate the applicability of the optimal designs for further analysis. The first one is obtained from a laboratory test to discover the performances and the efficiency of a given convenient design and the other presents the corresponding optimal designs for binary outcomes on determining suitable dosage levels which may best describe the response rates.

4.1. Example 1. Water treatment data

In the first example, we turn to a study on water purification treatment issues. In water purification plants, one of the first steps in a conventional water purification process is the addition of chemicals to assist in the removal of suspended particles in water. The particles, e.g. inorganic and organic, contribute to the turbidity and color of water. To study the effect of the chemical for removing existing contaminants in the water treatment process, an exploratory experiment is designed and performed. The experiment lasts for 100 days. For each day, the experimenter collects six samples of 500ml raw water from the fast mixing pool, records the original turbidity and PH level. In each sample, the experimenter adds different levels of chemical treatment respectively, and records the turbidity level after the treatment.

There are five variables involved: the original turbidity, dosage of coagulant, PH level, the turbidity after treatment and the elimination rate. The experimenter is interested in the connection among the variables of the coagulant dosage, the PH and the elimination rate. However in this dataset, there is only a weak correlation between the elimination rate and PH variables. Moreover, one of the main interests is to know the minimum dosage amount needed for effective water purification to a certain quality assurance level. To this end, we consider depicting the effectiveness of the water purification treatment with the simplex dispersion model. The experimental data to be modeled is presented in Fig. 3, where obvious relationships among the original turbidity, the coagulant dosage and the elimination rate are exhibited.

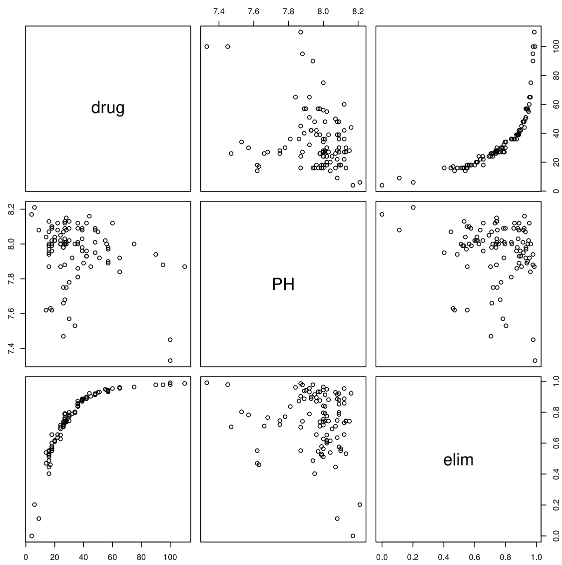

Fig. 3. The scatter plots of the water treatment data with three variables including coagulant dosage, PH level, and the elimination rate.

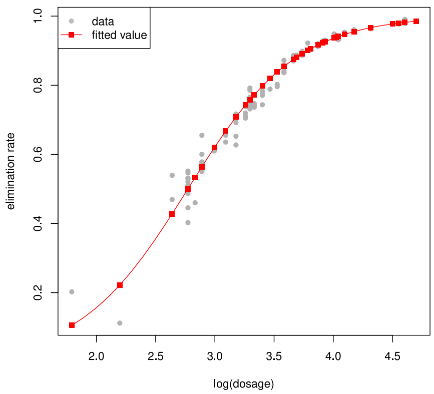

Fig. 4. The estimated mean curves for the water treatment data with the Simplex dispersion model (■) and the original elimination rate data is symbolized by "•".

The utilized response and explanatory variables are the elimination rate and the minimum log dosage. The maximum likelihood estimates (MLEs) of $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \sigma)$ for the water treatment data are $(-6.048, 2.182, 0.716)$, which can be obtained by solving the score equations in Song (2007). This estimative procedure can be carried out by the "VGAM" package in R. The estimated mean curves are shown in Fig. 4 which clearly indicates that a SD model is quite suitable for describing the relationship between the two variables under investigation. Moreover, from the design point of view, in future water purification quality assurance monitoring plans, the experimenter need only use the corresponding D-optimal designs, namely $\bar{\xi}_b$, for regular inspection of the effectiveness of the water purification treatment, and there is no need to use the usual uniform type of design with more than five supports which has rather low efficiency, as only a few extreme drug dosages would be needed in this application.

4.2. Example 2. Mortality data

In drug development, the relationship between the dose and a toxic response to a drug is important. A drug trail with varied levels of active ingredient was conducted to assess its toxicity on RatRiddance. Eleven different concentrations are tested, with about the same number of rats in each treatment. Table 3 denotes the concentration of the dosage and the proportion of the rats that did not survive (the mortality rate) for each dose (Peck et al. (2011), Chap. 5). The original data consisted of around 5000 observations; for each individual rat there was a (dose, response) pair, where the response was recorded as survived or not survived.

The maximum likelihood estimates (MLEs) of $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \sigma)$ are $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\sigma}) = (-1.399, 0.021, 0.678)$ for the mortality data, which can also be carried out by the "VGAM" package in R described in Section 4.2. The estimated mean curves are

Table 3
The experimental results of the mortality data for RatRiddance.

Concentration	20	40	60	80	100	120	140	160	180	200	240
Mortality rate	.225	.236	.398	.628	.678	.795	.853	.860	.921	.940	.968

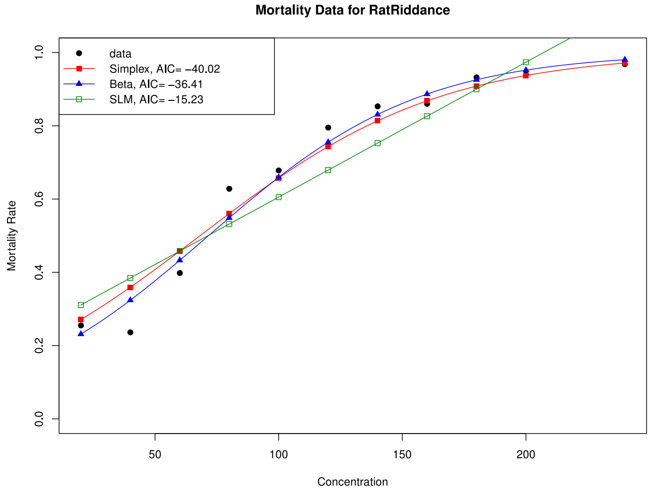

Fig. 5. The estimated mean curves for the Simplex dispersion model (■), the beta regression model (▲) and the simple linear regression model (□) and the original data is symbolized by "•".

shown in Fig. 5. Here, the Akaike's information criterion (AIC) is used for model comparison and a smaller value of AIC indicates a better fit. The Simplex dispersion model obtains the minimum AIC among the three candidate models, namely, the simple linear model, the beta regression model (Ferrari and Cribari-Neto (2004)) and the SD model. The transformed symmetric range of the dose concentration level is $\mathcal{X}_b = [-2.310, 2.310]$. Therefore, we can easily know that the D-optimal design for this data is $\bar{\xi}_b$ supported equally on the two ends of the concentration levels. The corresponding D-efficiency of the original experiment as in (16) is about 45.293%, which means that at least two replicates of the experiment are needed to achieve the same precision as that of the D-optimal design $\bar{\xi}_b$.

In the mortality response experiment, to be able to impose enough mortality on the litter critters for the study, sometimes a larger toxic agent would be needed. But we do not want to use higher toxic levels more than is necessary for experimenting due to safety and ecological concerns. The SD model would be useful for modeling such kinds of experiments.

5. Conclusion and discussion

In this study, the main purpose is to investigate the design problems of estimating unknown parameters under simplex dispersion models with continuous proportional data. The D-optimal designs of the simplex dispersion models with the linear logit link function for mean and a given dispersion parameter have been discussed. Due to the influence of the dispersion parameter and the selected design interval, the characteristics and results of the corresponding D-optimal designs are divided into four types of design patterns within the subclass of symmetric designs. In the present work, we derive results on the structure of optimal designs for the simplex dispersion model on particular symmetric design spaces. Two real examples are considered for illustrations of the benefits of using optimal designs with simplex dispersion models.

Determination of optimal designs analytically for estimation of unknown parameters in nonlinear models is interesting and challenging. Numerical programming can be adopted for optimal designs as a start for understanding the possible patterns of the optimal designs. However, with many different possible values of the parameters in the model, it would be helpful to provide the essential structures of the optimal designs theoretically with respect to the model parameters. Then we may have a better understanding of the behaviors of the optimal designs for simplex dispersion models and improve the efficiencies of experimental designs for proportional data.

In the case when the candidate designs cannot be made to be symmetric, in some cases it is possible to obtain a D-optimal design with three support points. For example, when $\sigma = 10$ and $b = 2.1$ where the four-point symmetric design is given in Table 2, a three-point D-optimal design can be found with a support not at the center as in

$$\xi^* = \begin{Bmatrix} 2.100 & 0.377 & -2.100 \\ 0.499 & 0.002 & 0.499 \end{Bmatrix}.$$

However, with the symmetric assumption, it is theoretically easier to characterize optimal designs.

The advantage of using the SD model is that the structure of the information matrix is simpler than that of the beta model. More explicitly, the dispersion parameter in the SD model can be estimated separately, but the parameters in the beta regression model do not have such a property. In this work, we consider the dispersion parameter as a known constant. The study of optimal designs where μ, σ^2 are both functions of the covariates is also worthy of future investigation.

As far as choices of the link function $g(\cdot)$, there are other possibilities such as $\log\mu$, $1/\mu$ or $-\log(-\log(\mu))$. Inevitably, the nice structure of the information matrix under the logit link may not hold for other links. How would the optimal designs change with respect to different links is again an interesting issue. Moreover here we have only investigated the optimal design problems in the simplex dispersion model with one variable. In Jørgensen and Lauritzen (2000), the multivariate dispersion case has also been illustrated. The design issue considered here can be extended to cases with multi-responses or multi-covariates. In summary, model robust design problems related to model link uncertainty, as well as design problems for multi-responses or multi-covariates data are all interesting problems worth studying with proportional data. Issues related to these problems will be studied further in the future.

Acknowledgments

The research of Prof. Hsu was supported in part by the Ministry of Science and Technology of Taiwan under grant MOST 105-2118-M-390-004. The research of Prof. Huang was supported in part by the Ministry of Science and Technology of Taiwan under grant MOST 103-2118-M-110-001-MY2. We would like to thank an Associate Editor and two anonymous referees for their insightful and constructive comments, which greatly improve the presentation of this paper.

Appendix A. Characterization of the weight functions and its properties

The derivative properties of $\lambda(z) = 3\tilde{\lambda}(z) = 3(\lambda_1(z)+\lambda_2(z))$, $\lambda_1(z)$, and $\lambda_2(z)$ in $M_\mu(\boldsymbol{\alpha}, \xi_{\boldsymbol{z}})$ are presented in Lemma A.2 for facilitating the study. For completeness, the properties about $\mu(z) = 1/(1+e^{-z})$ are summarized below.

Lemma A.1. *For $z \in \mathbb{R}$, the properties about $\mu(z)$ are*

(i) $\lim_{z\to-\infty}\mu(z) = 0$ *and* $\lim_{z\to\infty}\mu(z) = 1$.
(ii) $\mu(z)$ *is increasing and* $\mu(0) = 1/2$.
(iii) $0 < \mu(z) < 1$.
(iv) $0 < \mu(z)(1-\mu(z)) \leq 1/4$.

Lemma A.2. *The derivatives of the functions $\lambda_1(z)$, $\lambda_2(z)$ and $\tilde{\lambda}(z)$ are*

(i) $\dot{\lambda}_1(z) = \frac{d}{dz}\lambda_1(z) = (1-2\mu(z))\lambda_1(z)$.
(ii) $\dot{\lambda}_2(z) = \frac{d}{dz}\lambda_2(z) = -\left(c_\sigma\,(1-2\mu(z))\right)\Big/\lambda_1(z)$.
(iii) $\dot{\tilde{\lambda}}(z) = \frac{d}{dz}\tilde{\lambda}(z) = (1-2\mu(z))\,(h_\sigma(z)/\lambda_1(z))$, *where* $h_\sigma(z) = \lambda_1^2(z) - c_\sigma$.
(iv) $\ddot{\tilde{\lambda}}(z) = \frac{d^2}{dz^2}\tilde{\lambda}(z) = Q_1(z) + Q_2(z)$ *where* $Q_1(z) = -2h_\sigma(z)$, $Q_2(z) = (1-2\mu(z))^2\tilde{\lambda}(z)$.

Lemma A.2 can be easily obtained and the proof is omitted here.

The function $h_\sigma(z) = \lambda_1^2(z) - c_\sigma$, whose sign depends on the given σ and z_0 as presented in Lemma A.3, plays an important role for determining the patterns of the weight functions and the investigation of optimal designs.

Lemma A.3. *If $\sigma \leq \sigma_1^*$, $z \in \mathbb{R}$ or $\sigma > \sigma_1^*$ and $|z| \geq z_0$, then $h_\sigma(z) \leq 0$; Otherwise, $h_\sigma(z) > 0$.*

Proof. By taking derivative of $h_\sigma(z)$, we have

$$\frac{d}{dz}h_\sigma(z) = 2\lambda_1^2(z)(1-2\mu(z))$$

yielding that $h_\sigma(z)$ is strictly increasing (decreasing) for $z < 0$ ($z > 0$), and has a unique maximum at $z = 0$. Then when $\sigma \leq \sigma_1^*$, namely $c_\sigma = 1/3\sigma^2 \geq 1/16$, the first part can be easily obtained as

$$h_\sigma(z) \leq h_\sigma(0) = 1/16 - c_\sigma \leq 0,\ \forall z \in \mathbb{R}.$$

When $\sigma > \sigma_1^*$, $h_\sigma(0) = 1/16 - c_\sigma > 0$, and $\lim_{z\to\infty} h_\sigma(z) = -c_\sigma < 0$. Given that $h_\sigma(z)$ is strictly decreasing for $z > 0$, there exists a unique $z_0 > 0$ such that $h_\sigma(z_0) = 0$ as shown in (7). A similar argument yields that there exists a unique $z_0' < 0$ such that

$$z_0' = \arg\left\{z : h_\sigma(z) = 0,\ z < 0 \text{ and } \sigma > \sigma_1^*\right\} = -z_0.$$

From Lemma A.1 and (7), it is clear that $h_\sigma(z) \leq 0$ for all $z \in \mathbb{R}$ when $\sigma \leq \sigma_1^*$, or for $|z| \geq z_0$ when $\sigma > \sigma_1^*$. It is also obvious that $h_\sigma(z) > 0$ when $|z| < z_0$ and $\sigma > \sigma_1^*$ from the above discussions. ■

Now, we focus on the patterns of the first and second derivatives of $\lambda(z)$ and where the local extremes and inflection point of $\lambda(z)$ occur.

Lemma A.4. *For $\sigma > 0$ and z_0 as defined in (7),*

(i) if $\sigma \le \sigma_1^*$, the only critical point of $\lambda(z)$ is at $z = 0$, which is a minimum;
(ii) if $\sigma > \sigma_1^*$, then the critical points of $\lambda(z)$ are at $z = 0$ and $\pm z_0$, which are the local maximum and minimums, respectively. Moreover, there exists a unique inflection point in $\lambda(z)$ for $z \in (0, z_0)$, and $\lim_{z\to\pm\infty} \lambda(z) = \infty$.

Proof.

(i) For $\sigma \le \sigma_1^*$, note that by Lemma A.2(iii),(iv), we have

$$\dot{\lambda}(z) = 3(1 - 2\mu(z))(h_\sigma(z)/\lambda_1(z)),\ \ \ddot{\lambda}(z) = 3(Q_1(z) + Q_2(z)).$$

Then by Lemma A.3, it can be confirmed easily that

$$\dot{\lambda}(z) < 0 (> 0) \text{ for } z < (>) 0;\ \ \ddot{\lambda}(z) > 0,\ \ \forall z \in \mathbb{R};$$

and $\dot{\lambda}(0) = 0$, $\ddot{\lambda}(0) > 0$. It is clear that $z = 0$ is the only critical point of $\lambda(z)$, which is a minimum.

(ii) Again by Lemma A.2(iii), (iv) and Lemma A.3, if $\sigma > \sigma_1^*$, it can be confirmed with ease

(1) $\dot{\lambda}(z) > 0$ for $z \in (-z_0, 0) \cup (z_0, \infty)$; $\dot{\lambda}(z) < 0$, $z \in (-\infty, -z_0) \cup (0, z_0)$; $\dot{\lambda}(0) = \dot{\lambda}(\pm z_0) = 0$;
(2) $\ddot{\lambda}(z) > 0, \forall |z| > z_0$; $\ddot{\lambda}(0) > 0$;
(3) there exists a unique inflection point in $\lambda(z)$ for $z \in (0, z_0)$.

Then $z = 0$ and $\pm z_0$, which are the local maximum and minimums, respectively. According to Lemma A.2(iv) and for $c_\sigma < 1/16$,

$$Q_1(0) = -2(1/16 - c_\sigma) < 0,\ \text{and}\ Q_2(0) = 0.$$

This yields $\ddot{\lambda}(0) < 0$ and $\ddot{\lambda}(z_0) > 0$. Hence there exists at least one $z^* \in (0, z_0)$ such that $\ddot{\lambda}(z^*) = 0$, which implies that $z = z^*$ is an inflection point of $\lambda(z)$. Moreover, since $\dot{\lambda}(z) < 0$, $\forall z \in (0, z_0)$, it follows that there does not exist any critical point in $(0, z_0)$ and $z = 0$, z_0 are the corresponding locally maximum and minimum points of $\lambda(z)$. Therefore, z^* is the unique inflection point in $(0, z_0)$. ■

Lemma A.5. *For σ_1^*, z_0 and $z \in \mathbb{R}^+$, some useful properties for $L(z) = \dot{\lambda}(z)\big/\lambda(z)$ are described as follows.*

(i) For $\sigma \le \sigma_1^*$, $L(z) > 0$, $\forall z > 0$;
(ii) For $\sigma > \sigma_1^*$ and $z \ge z_0$, $L(z) > 0$;
(iii) $\lim_{z\to\infty} L(z) = 1$.

Lemma A.5 can be easily obtained by the assist of Lemmas A.1∼ A.4 and the proof is omitted here.

Appendix B. Proofs of Lemmas 1, 2 and Proposition 1

B.1. Proof of Lemma 1

From the equivalence theorem, the support points of an optimal design should be among those with maximum values of the corresponding dispersion function. By taking the derivative of $d(z, \bar{\xi})$ with respect to z, we are able to determine where the possible position of the local maximum of the dispersion function is within the design region $\mathcal{X}_{z,b}$ for design $\bar{\xi}$. Note that since the dispersion function for a symmetric design is a symmetric function, we only need to discuss the pattern of $d(z, \bar{\xi})$ for $z \ge 0$.

Determining the roots of the derivative of $d(z, \bar{\xi})$ with respect to z, requires solving the following equation within $\mathcal{X}_b$

$$\frac{\dot{\lambda}(z)}{\lambda(z)} = -\frac{2z}{\kappa(\bar{\xi}) + z^2},$$

where $L(\lambda) = \dot{\lambda}(z)/\lambda(z)$, and $\kappa(\bar{\xi}) = \kappa_1(\bar{\xi})/\kappa_2(\bar{\xi})$ is a constant depending on the design. We now return to verify the problem of finding an interior point $a \in [-b, b]$, satisfying (11). To find all possible a is equivalent to finding the points of intersection in $[0, b)$ between $L(z)$ with the function on the right hand side of (11). According to Lemma A.5(i), (ii), $L(z)$ cannot be equal to $-2z/(\kappa(\bar{\xi}) + z^2)$ when $\sigma \le \sigma_1^*$ and if the interior point a exists, it must be constrained within $(0, z_0)$ when $\sigma > \sigma_1^*$. Since for a symmetric design $\bar{\xi}$ and positive $\kappa(\bar{\xi})$,

$$\lim_{z\to\pm\infty} d(z, \bar{\xi}) = \infty, \quad \lim_{z\to\infty} L(z) = 1,\ \text{and}\ \lim_{z\to\infty} \frac{2z}{\kappa(\bar{\xi}) + z^2} = 0, \tag{B.1}$$

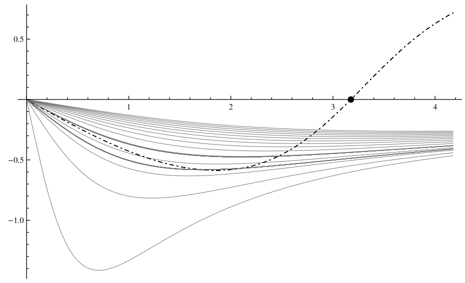

Fig. B.6. The fluctuation of the functions on the two sides of (11) when $\sigma = 15$. The dashed line is the curve of the left hand side function in (11) and other lines under the line of $y = 0$ are the curves of the right hand side function in (11) with different κ. The value of z_0 is labeled as the solid dot.

it reveals that functions on the two sides of (11) can have only one local extremum for $z > 0$. These properties yield that there are at most two intersection points between $L(z)$ and $2z/(\kappa(\bar{\xi}) + z^2)$ when $z \in [0, z_0]$, as exhibited in Fig. B.6 for illustrations.

B.2. Proof of Lemma 2

If the coefficient of γ^2 in $g(a, \gamma)$, defined as (12) is positive for a given a, or equivalently if either of the following inequalities holds

$$\lambda(b)/\lambda(a) > 1, \quad \text{or} \quad \lambda(b)/\lambda(a) \le a^2/b^2, \tag{B.2}$$

the maximum of $g(a, \gamma)$ will occur at either $\gamma = 0$, or $\gamma = 1$. It is easy to find $\gamma_{\max}$, the location of the maximum of $g(a, \gamma)$, by taking partial derivative of $g(a, \gamma)$ with respect to γ and solving $\partial g(a, \gamma)/\partial\gamma$=0. We have

$$\gamma_{\max} = \frac{2\lambda^2(b)b^2 - \lambda(a)\lambda(b)(a^2 + b^2)}{2\left(\lambda(b) - \lambda(a)\right)\left(\lambda(b)b^2 - \lambda(a)a^2\right)}. \tag{B.3}$$

Then according to the value of $\gamma_{\max}$, we can argue that the following assertions in (i)(a) to (c) hold.

(i) For $\lambda(b)/\lambda(a) > 1$ or $\left(a^2 + b^2\right)/\left(2b^2\right) \le \lambda(b)/\lambda(a) \le 1$, we have that $\gamma_{max} < 0$. Therefore within $0 \le \gamma \le 1$, $\gamma_a^* = 0$.
(ii) Similarly, when $\lambda(b)/\lambda(a) \le a^2/b^2$ or $a^2/b^2 < \lambda(b)/\lambda(a) \le 2a^2/\left(a^2 + b^2\right)$, we have that $\gamma_{\max} > 1$ and $\gamma_a^* = 1$.
(iii) Then when $2a^2/\left(a^2 + b^2\right) < \lambda(b)/\lambda(a) < \left(a^2 + b^2\right)/\left(2b^2\right)$, it can be seen that $0 < \gamma_a^* < 1$.

When $a = 0$, $g(a, \gamma)$ can be simplified as follows

$$g(0, \gamma) = \lambda(b)b^2\left[(\lambda(b) - \lambda(0))\gamma^2 + (\lambda(0) - 2\lambda(b))\gamma + \lambda(b)\right].$$

Similar arguments yield that the assertions in (ii) hold when a equals 0.

B.3. Proof of Proposition 1

Assertions in (i) and (ii)(a) hold according to Lemmas 2, A.2, A.4 and arguments based on values of σ_1^*, and σ_2^* and b^*.

(ii)(b) According to Lemma A.4, there are two intersections, denoted as z_1 and z_2, of the two functions $\lambda(z)$, $2\lambda(0)$, $\forall z \in \mathbb{R}^+$ where

$$z_1 = \log\left[\frac{1}{32}\left(-16 + 3\sigma^2 - P_1 + \sqrt{18\sigma^4 + 32(-16 + P_1) - 6\sigma^2(128 + P_1)}\right)\right]; \tag{B.4}$$

$$z_2 = \log\left[\frac{1}{32}\left(-16 + 3\sigma^2 + P_1 + \sqrt{-32(16 + P_1) + 6\sigma^2(-128 + 3\sigma^2 + P_1)}\right)\right], \tag{B.5}$$

and $P_1 = \sqrt{256 - 672\sigma^2 + 9\sigma^4}$.

(ii)(c) Firstly, it is known that $\lambda(b) > \lambda(z)$ when $b > b^*$. It can be established easily that $\lambda(b)b - \lambda(z)z > 0$. When $z_2 \le b < b^*$, there exists a $z_b \in [0, z_1]$ satisfying

$$z_b = \arg\{z : \lambda(z) = \lambda(b), z \in [0, z_1],\ z_2 < b < b^* \text{ and } \sigma > \sigma_2^*\}.$$

For $z \in [z_b, b]$, we have $\lambda(b) \ge \lambda(z)$ and

$$\begin{aligned}\lambda(b)b - \lambda(z)z &= \lambda(b)b - \lambda(z)b + \lambda(z)b - \lambda(z)z \\ &= b(\lambda(b) - \lambda(z)) + \lambda(z)(b - z) > 0.\end{aligned}$$

Otherwise when $z \in [0, z_b)$, it can be verified that

$$\lambda(b)b - \lambda(z)z > \lambda(b)b - \lambda(z_b)z_b = \lambda(b)(b - z_b) > 0.$$

Therefore, this assertion holds.

Appendix C. Proof of Theorems 1 and 2

C.1. Proof of Theorem 1

(i) For $\{0 < \sigma \leq \sigma_2^*\}$, according to Proposition 1(i) and when $\sigma < \sigma_1^*$ or $\sigma \geq \sigma_1^*$ then with Lemma 2(i)(a) or (ii), it yields that $\gamma_a^* = 0$ or $\gamma_0^* = 0$, respectively.

(ii) For $\{\sigma > \sigma_2^*\}$,
(a) when $b \leq z_1$, with Proposition 1(ii)(c) and Lemma 2(i)(a) or (b), it can be shown that $\gamma_a^* = 0$ or $\gamma_a^* = 1$.
(b) when $b \in (z_1, z_2)$, with Proposition 1(ii)(c) and Lemma 2(ii), it yields that $\gamma_0^* \neq 0$.
(c) when $b \geq z_2$, with Proposition 1(ii)(c) and Lemma 2(ii), it yields that $\gamma_0^* = 0$.

Then results of Theorem 1 are obtained and the proofs are completed.

C.2. Proof of Theorem 2

It has been confirmed that the dispersion function for a symmetric design is a symmetric function in Section 3. So here, we only need to discuss cases when $z \geq 0$.

(i)(a) For $0 < \sigma \leq \sigma_2^*$, we have $d(z, \bar{\xi}_b^2) = (b^2 + z^2)/b^2 \times \lambda_b(z)$, and

$$\frac{\partial}{\partial z} d(z, \bar{\xi}_b^2) = \frac{z}{b^2 \lambda(b)} \left(2\lambda(z) + z\dot{\lambda}(z) \right).$$

By Lemma A.4, $\frac{\partial}{\partial z} d(z, \bar{\xi}_b^2)|_{z=0} = 0$ and $\frac{\partial}{\partial z} d(z, \bar{\xi}_b^2) > 0$ for $z \in [z_0, \infty)$. Otherwise, we know that $\lambda(z) > \lambda(0)/2$ for $0 < \sigma \leq \sigma_2^*$ and it can be verified that for $z \in (0, z_0)$

$$\begin{aligned} \left(2\lambda(z) + z\dot{\lambda}(z) \right) &> \lambda(0) + z\,(1 - 2\mu(z))\,(h_\sigma(z)/\lambda_1(z)) \\ &= \lambda(0) - z\,(1 - 2\mu(z))\, c_\sigma + z\,(1 - 2\mu(z))\, \lambda_1(z) \end{aligned} \tag{C.1}$$

where the value of the first two parts in (C.1) is positive. Moreover in (C.1), $z\,(1 - 2\mu(z))\,\lambda_1(z)$ does not depend on σ and its minimum value equals $-0.160 > -\lambda(0)(= -\frac{1}{4} - \frac{4}{3\sigma^2})$. This yields that the derivative of $d(z, \bar{\xi}_b^2)$ is greater than 0. Hence, $d(z, \bar{\xi}_b^2)$ is strictly increasing for $z > 0$ and the maximum value of $d(z, \bar{\xi}_b^2)$ occurs at $z = b$. Consequently, $d(z, \bar{\xi}_b^2) \leq 2$ and the design $\bar{\xi}_b^2$ is D-optimal according to the equivalence theorem.

(i)(b) For the other case with $\sigma > \sigma_2^*$ and $b \geq z_2$, we provide a proof for the particular setting $b = z_2$ and then study the performance of the design $\bar{\xi}_b^2$ for $b > z_2$. Moreover, if we can verify that $\lambda(b)/\lambda(z)$ is greater than $(z^2 + b^2)/(2b^2)$ in this case, then, with $z \in [0, b]$, $\bar{\xi}_b^2$ can be shown to be D-optimal by Lemma 2. The above argument may be verified from observing the patterns of the two functions, $\lambda(b)/\lambda(z)$ and $(z^2 + b^2)/(2b^2)$.
For $\sigma > \sigma_2^*$ and $b = z_2$, the first and second derivatives of $\lambda(b)/\lambda(z)$ with respect to z are

$$\begin{aligned} \frac{\partial}{\partial z}\left(\frac{\lambda(b)}{\lambda(z)}\right) &= -\frac{\lambda(b)}{\lambda^2(z)}\dot{\lambda}(z); \\ \frac{\partial^2}{\partial z^2}\left(\frac{\lambda(b)}{\lambda(z)}\right) &= -\frac{\lambda(b)}{\lambda^2(z)}\ddot{\lambda}(z) + 2\frac{\lambda(b)}{\lambda^3(z)}\left(\dot{\lambda}(z)\right)^2 \end{aligned} \tag{C.2}$$

By Lemma A.4, it can be obtained that (C.2) equals 0 at $z = 0, z_0$, which are the local minimum and maximum values, respectively. It is easy to show that $(z^2 + b^2)/(2b^2)$ is increasing in $z \in [0, b]$. One more characteristic is that $\lambda(b)/\lambda(z) - (z^2 + b^2)/(2b^2)$ equals 0 when $z = 0, b$. Hence, there are two possible patterns for $\lambda(b)/\lambda(z)$ and $(z^2+b^2)/(2b^2)$; one has an intersection between the two functions in $(0, b)$; and the other is that $\lambda(b)/\lambda(z)$ is greater than $(z^2 + b^2)/(2b^2)$. The curve of $\lambda(b)/\lambda(z) - (z^2 + b^2)/(2b^2)$ is similar to $\lambda(b)/\lambda(z)$ and its minimum value is 0 at $z = 0$ when $b = z_2$. This indicates that only the case $\lambda(b)/\lambda(z)$ greater than $(z^2 + b^2)/(2b^2)$ holds. Therefore, the argument that design $\bar{\xi}_b^2$ is D-optimal with $\sigma > \sigma_2^*, z \in [-b, b]$ and $b = z_2$ holds by Lemma 2.
Now, for fixed $z \in [-b, b]$, $\lambda(b)/\lambda(z)$ is increasing and $(z^2 + b^2)/(2b^2)$ is decreasing when b is increasing from z_2 to ∞. The difference between $\lambda(b)/\lambda(z)$ and $(z^2 + b^2)/(2b^2)$ becomes larger when b increases. It yields that $\sigma > \sigma_2^*, z \in [-b, b]$ and $b \geq z_2$ holds through Lemma 2.

Similarly, the cases of (ii) and (iii) can also be derived by the results of Proposition 1 and Lemma 2.

References

Chang, F.C., Lin, G.C., 1997. *D*-optimal designs for weighted polynomial regression. J. Statist. Plann. Inference 62, 317–331.
Dette, H., Haines, L.M., Imhof, L., 1999. Optimal designs for rational models and weighted polynomial regression. Ann. Statist. 27, 1272–12193.
Dette, H., Trampisch, M., 2010. A general approach to *D*-optimal designs for weighted univariate polynomial regression models. J. Korean Stat. Soc. 39, 1–26.
Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31 (7), 799–815.
Ford, I., Torsney, B., Wu, C.F.J., 1992. The use of a canonical form in the construction of locally optimal designs for non-linear problems. J. R. Stat. Soc. Ser. B Stat. Methodol. 54, 569–583.
Giovagnoli, A., Pukelsheim, F., Wynn, H., 1987. Group invariant orderings and experimental designs. J. Statist. Plann. Inferenece 17, 159–171.
Huang, M.-N.L., Chang, F.C., Wong, W.K., 1995. *D*-optimal designs for polynomial regression without an intercept. Statist. Sinica 5, 441–458.
Imhof, L., Krafft, O., Schaefer, M., 1988. *D*-optimal designs for polynomial regression with weight function $x/(1+x)$. Statist. Sinica 8, 1271–1274.
Jørgensen, B., 1997. The Theory of Dispersion Models. Chapman and Hall, London.
Jørgensen, B., Lauritzen, S.L., 2000. Multivariate dispersion models. J. Multivariate Anal. 74, 267–281.
Karlin, S., Studden, W.J., 1966. Optimal experimental designs. Ann. Math. Stat. 37, 783–815.
Kiefer, J., 1959. Optimum experimental designs. J. R. Stat. Soc. Ser. B Stat. Methodol. 21, 272–319.
Kiefer, J., 1961. Optimal designs in regression problems, II. Ann. Math. Stat. 32, 298–325.
Latif, S., Zafar Yab, M., 2015. *D*-optimal designs for beta regression models with single predictor. J. Stat. Comput. Simul. 85 (9), 1709–1724.
Minkin, S., 1987. Optimal designs for binary data. J. Amer. Statist. Assoc. 82, 1098–1103.
Musrati, A.K., 1992. Locally Optimal Designs for Binary and Weighted Regression Models (Ph.D. thesis). ProQuest Dissertations & Theses.
Peck, R., Olsen, C., Devore, J., 2011. Introduction to Statistics and Data Analysis. Cengage Learning, URL https://books.google.com.tw/books?id=AsdV9duleTsC.
Qiu, Z., Song, P.-K., Tan, M., 2008. Simplex mixed-effects models for longitudinal proportional data. Scand. J. Stat. 35, 577–596.
Silvey, S.D., 1980. Optimal Design: An Introduction to the Theory for Parameter Estimation. Chapman and Hall, London.
Sitter, R.R., Torsney, B., 1995a. *D*-optimal designs for generalized linear models. In: MODA4—Advances in Model-Oriented Data Analysis. Springer, pp. 87–102.
Sitter, R.R., Torsney, B., 1995b. Optimal designs for binary response experiments with two design variables. Statist. Sinica 5, 405–419.
Sitter, R.R., Wu, C.F.J., 1993. Optimal designs for bnary response experiments: fieller, d and a criteria. Scand. J. Stat. 20, 329–341.
Song, P.-K., 2007. Correlated Data Analysis: Modeling, Analytics and Applications. Springer, New York.
Song, P.-K., Tan, M., 2000. Marginal models for longitudinal continuous proportional data. Biometrics 56, 496–502.
Torsney, B., Musrati, A., 1993. On the construction of optimal designs with applications to binary response and to weighted regression models. Model-Oriented Data Anal. 37–52.
White, L.V., 1973. An extension of the general equivalence theorem to nonlinear models. Biometrika 60, 345–348.
Whittle, P., 1973. Some general points in the theory of optimal experimental design. J. R. Stat. Soc. Ser. B Stat. Methodol. 35, 123–130.
Wu, Y., Fedorov, V., Propert, K., 2005. Optimal design for dose response using beta distributed responses. J. Biopharm. Statist. 15, 753–771.
Yang, M., Zhang, B., Huang, S., 2011. Optimal designs for generalized linear models with multiple design variables. Statist. Sinica 21, 1415–1430.