



D-optimal and nearly *D*-optimal exact designs for binary response on the ball

Martin Radloff¹ · Rainer Schwabe¹

Received: 6 January 2023 / Revised: 28 February 2023 / Published online: 27 April 2023
© The Author(s) 2023

Abstract

In this paper the results of Radloff and Schwabe (Stat Papers 60:165–177, 2019) will be extended for a special class of symmetrical intensity functions. This includes binary response models with logit and probit link. To evaluate the position and the weights of the two non-degenerated orbits on the k -dimensional ball usually a system of three equations has to be solved. The symmetry allows to reduce this system to a single equation. As a further result, the number of support points can be reduced to the minimal number. These minimally supported designs are highly efficient. The results can be generalized to arbitrary ellipsoidal design regions.

Keywords Binary response model · *D*-optimality · k -dimensional ball · Logit and probit model · Multiple regression model · Simplex

Mathematics Subject Classification 62K05 · 62J12

1 Introduction

Response surface methodology is often used in engineering experiments to describe the effect of various factors of influence (explanatory variables) on the outcome of a technical system. There, some statistical model is assumed to be valid in a vicinity of a target setting for the explanatory variables. Depending on the shape of the experimental design used, this vicinity covers a cubical region for factorial designs or a spherical region for (central) composite designs where additional axial points are included which

✉ Martin Radloff
martin.radloff@ovgu.de

Rainer Schwabe
rainer.schwabe@ovgu.de

¹ Institute for Mathematical Stochastics, Otto-von-Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany

have the same distance to the target setting as the factorial points on the cube, see e. g. Box and Draper (1987).

For linear models, these designs share some nice features like rotatability in the situation of multiple linear regression which makes it reasonable to use a spherical design region from which the experimental settings may be chosen. There, optimality properties were obtained in early work by Kiefer (1961) and Farrell et al. (1967) which discuss polynomial regression on the ball. These ideas were followed up by papers in which also only linear problems were focused on. So Lau (1988) fitted polynomials on the k -dimensional unit ball by using canonical moments. In Dette et al. (2005, 2007) and Hirao et al. (2015) harmonic polynomials and Zernike polynomials were used to be fitted on the unit disc (two-dimensional), the three-, and the k -dimensional unit ball.

On the other hand generalized linear models are well-examined and used in practical application, in particular, if the response is binary or consists of count data. Optimal design in the case of logit or probit models, for example, have been investigated by Ford et al. (1992) and Biedermann et al. (2006) on an interval which may be considered as a “ball” in one dimension. But there does not seem to be much literature available which extends these results to proper higher dimensional spherical regions. In Radloff and Schwabe (2019a) we made a first attempt to bring non-linearity or, more specifically, generalized linear models and spherical design regions together in the context of design optimization. The results therein were extended to a wider class of non-linear models in Radloff and Schwabe (2019b).

In the present paper, we will start with the model description in Sect. 2 and give a brief overview of the findings so far in Sect. 3. Then we will consider a special class of intensity functions which allows to reduce the complexity of finding (locally) D -optimal designs in Sect. 4. Thereafter we will tackle the problem, that the optimal designs are not exact designs in general, by establishing highly efficient designs on the ball in Sect. 5. Some basic notation and some proofs are given in Appendix A and Appendix B, respectively.

2 General model description

The outcome Y of an experiment may be influenced by a set $\mathbf{x} = (x_1, \dots, x_k)^\top$ of k explanatory variables x_1, \dots, x_k , $k \geq 1$, such that the distribution of a single response Y_i is determined by the corresponding experimental setting \mathbf{x}_i . In particular, the mean response $E(Y_i) = h(\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta})$ is a one-to-one function h of the linear predictor $\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}$, where \mathbf{f} is a p -dimensional vector of regression functions, $p \geq k$, and $\boldsymbol{\beta}$ is a p -dimensional vector of parameters. While the functions h and \mathbf{f} are assumed to be known, statistical inference is to be made on the parameter vector $\boldsymbol{\beta}$. In particular, in a linear model, the function h is the identity while, for generalized linear models, h is the inverse link function. However, the function h may be more general, e. g. in models with censored observations.

Under distributional assumptions on the response Y_i , the influence of the corresponding experimental setting \mathbf{x}_i on the performance of the statistical inference may be measured by the elemental information matrix $\mathbf{M}(\mathbf{x}_i, \boldsymbol{\beta})$. In generalized linear

models, also the variance $\text{Var}(Y_i) = \sigma^2(\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta})$ of the response Y_i is a function of the linear predictor only, and the elemental information matrix can be represented as

$$\mathbf{M}(\mathbf{x}_i, \boldsymbol{\beta}) = \lambda\left(\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}\right) \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^\top, \tag{1}$$

where the intensity function λ is given by $\lambda(z) = h^2(z)/\sigma^2(z)$ such that the intensity $\lambda(\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta})$ only depends on the linear predictor $\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}$. In a linear model, the intensity function λ is constant. But, also in other situations, the elemental information may be of the form (1) with suitable intensity function λ like for censored data. Thus, we will suppose throughout in the following that the elemental information matrix has the form (1).

Under the assumption of independent observations Y_1, \dots, Y_n at experimental settings $\mathbf{x}_1, \dots, \mathbf{x}_n$, the information matrix $\mathbf{M}((\mathbf{x}_1, \dots, \mathbf{x}_n), \boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{M}(\mathbf{x}_i, \boldsymbol{\beta})$ of the whole experiment is given by the sum of the elemental information matrices at the single settings. The collection $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ of the experimental settings will be called an (exact) design. The performance of a design can be measured in terms of the information matrix because the maximum-likelihood estimator of the parameter vector $\boldsymbol{\beta}$ is asymptotically normal with (asymptotic) covariance matrix proportional to the inverse of the information matrix under mild regularity conditions. The aim of design optimization is then to find experimental settings $\mathbf{x}_1, \dots, \mathbf{x}_n$ from a design region \mathcal{X} of potential settings which maximize the information in a certain sense. Here, we will make use of the D -criterion which is most popular in applications and which aims at maximizing the determinant of the information matrix. In terms of the (asymptotic) covariance matrix, the D -criterion can be interpreted as minimization of the volume of the (asymptotic) confidence ellipsoid for the whole parameter vector $\boldsymbol{\beta}$. Note that the information matrix depends on the value of the parameter vector $\boldsymbol{\beta}$. Hence, also the optimal design will depend on $\boldsymbol{\beta}$, and we will consider local D -optimality at some prespecified $\boldsymbol{\beta}^0$.

As in Radloff and Schwabe (2019a) and Radloff and Schwabe (2019b), where we described (locally) D -optimal designs for two special classes of linear and non-linear models, we consider as the design region \mathcal{X} the k -dimensional unit ball $\mathbb{B}_k = \{\mathbf{x} \in \mathbb{R}^k : x_1^2 + \dots + x_k^2 \leq 1\}$ and a multiple regression model for the linear predictor

$$\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

with regression function $\mathbf{f} : \mathbf{x} \mapsto (1, x_1, \dots, x_k)^\top$, and parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^\top \in \mathbb{R}^{k+1}$ such that the dimension of the parameter vector is $p = k + 1$. Further, we relax the concept of (exact) designs to the class of (generalized) designs ξ in the spirit of Kiefer (1959). Here, a generalized design means an arbitrary probability measure on the design region $\mathcal{X} = \mathbb{B}_k$ which is not necessarily discrete, as commonly assumed in the literature on approximate design theory, but may be continuous. The standardized information matrix of a (generalized) design ξ is then defined as

$$\mathbf{M}(\xi, \boldsymbol{\beta}) = \int \mathbf{M}(\mathbf{x}, \boldsymbol{\beta}) \xi(d\mathbf{x}) = \int \lambda\left(\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}\right) \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^\top \xi(d\mathbf{x}),$$

which reduces to $\frac{1}{n}M((x_1, \dots, x_n), \beta)$ in the case of a discrete design associated with the (exact) design (x_1, \dots, x_n) . Then, concerning (local) D -optimality, a design $\xi_{\beta^0}^*$ with regular information matrix $M(\xi_{\beta^0}^*, \beta^0)$ is (locally) D -optimal (at β^0) if $\det(M(\xi_{\beta^0}^*, \beta^0)) \geq \det(M(\xi, \beta^0))$ for all probability measures ξ on the design region $\mathcal{X} = \mathbb{B}_k$.

3 Prior results

In Radloff and Schwabe (2016) we stated results on equivariance and invariance in models, where the elemental information matrix is of the form (1). By rotating the design region \mathbb{B}_k and the parameter space \mathbb{R}^{k+1} simultaneously by $\mathbf{g} : \mathbf{x} \mapsto (g_1(\mathbf{x}), \dots, g_k(\mathbf{x}))^\top$ such that $g_1(\mathbf{x})$ points into the direction of the maximum value $\max_{\mathbf{x} \in \mathbb{B}_k} (\mathbf{f}(\mathbf{x})^\top \beta)$ of the linear predictor on the ball, $\tilde{\mathbf{g}} : \beta \mapsto (\beta_0, \sqrt{\beta_1^2 + \dots + \beta_k^2}, 0, \dots, 0)^\top$, and this reparameterization leaves the intensity $\lambda(\mathbf{f}(\mathbf{g}(\mathbf{x}))^\top \tilde{\mathbf{g}}(\beta)) = \lambda(\mathbf{f}(\mathbf{x})^\top \beta)$ unchanged. Design optimality carries over from one parameterization to the other by the transformation \mathbf{g} on the design region or its inverse \mathbf{g}^{-1} , respectively.

Thus, we can confine our investigations to parameter vectors of the form

$$\beta^0 = (\beta_0, \beta_1, 0, \dots, 0)^\top \quad (2)$$

with $\beta_1 \geq 0$ in which, apart from the intercept term β_0 , only the slope β_1 for the component x_1 may differ from zero. A (locally) D -optimal design $\xi_{\beta^0}^*$ obtained for β^0 of the form (2) then yields a (locally) D -optimal design $\mathbf{g}^{-1}(\xi_{\beta^0}^*)$ for a general β^0 , where $\mathbf{g}^{-1}(\xi_{\beta^0}^*)$ is the (measure-theoretic) image of $\xi_{\beta^0}^*$ under the mapping \mathbf{g}^{-1} .

For β^0 of the form (2), the linear predictor reduces to

$$\mathbf{f}(\mathbf{x})^\top \beta^0 = \beta_0 + \beta_1 x_1 \quad (3)$$

where $\beta_1 \geq 0$. Note that the linear predictor and, thus, the intensity only vary in x_1 , while these quantities are constant in the other directions orthogonal to the direction of x_1 .

If $\beta_1 = 0$, the linear predictor and the intensity function will be constant. This results in a (locally) D -optimal design which does not depend on β^0 and is the same as in the corresponding linear model. According to Pukelsheim (1993, Sect. 15.12) such an optimal design consists of equally weighted vertices of a regular simplex inscribed in the unit sphere, which is the boundary of the design region, and the orientation of the simplex may be chosen arbitrarily. So we only need to consider $\beta_1 > 0$ from now on.

For β^0 of the form (2), the (local) D -criterion is rotationally invariant with fixed first component x_1 , i. e. invariant with respect to the subgroup of all orthogonal transformations in the orthogonal group $O(k)$ which leave the x_1 -component unchanged. Then,

there will be a (locally) D -optimal (generalized) design $\xi_{\beta^0}^*$ which is also rotationally invariant with fixed x_1 .

If we regard a design ξ on \mathcal{X} as a joint distribution of the projections onto the components of \mathbf{x} , then it can be decomposed into a marginal design (marginal distribution) ξ_1 on the first component x_1 supported on the marginal design region \mathcal{X}_1 , which is the projection of \mathcal{X} onto x_1 , and a probability kernel η which, for every x_1 , provides a conditional design $\eta(x_1, \cdot)$ on the conditional design region $\mathcal{X}_2(x_1)$, which is the x_1 -cut of \mathcal{X} , such that $\xi = \xi_1 \otimes \eta$, where “ \otimes ” denotes the measure-theoretic product.

In the present case of a (locally) D -optimal rotationally invariant design ξ^* , the conditional design $\bar{\eta}(x_1, \cdot)$ is the uniform distribution on the surface of a $(k - 1)$ -dimensional ball with radius $\sqrt{1 - x_1^2}$ —the outmost orbit at position x_1 . The D -optimal design is then of the form $\xi^* = \xi_1^* \otimes \bar{\eta}$. As a consequence of the decomposition, the multidimensional problem reduces to a one-dimensional marginal problem. Only the marginal design ξ_1 has to be optimized, i.e. the positions x_1 of the orbits and their weights have to be determined. To finally get an exact design, the uniform orbits have to be discretized, for example, by using regular simplices.

In Radloff and Schwabe (2019a) we started with models where the intensity function belongs to the *class of monotonic functions*. This means the first derivative of the intensity function λ' is positive (or negative) on \mathbb{R} . Such models satisfying four particular conditions on the intensity function λ similar to (A1) to (A4) below have been investigated in one dimension, for example, by Konstantinou et al. (2014) and on multidimensional cuboids or orthants by Schmidt and Schwabe (2017). The results therein can be applied, for example, to Poisson regression and negative binomial regression as well as special proportional hazard models with censoring, see Schmidt and Schwabe (2017).

In Radloff and Schwabe (2019b) two of the four conditions were modified to (A2') and (A3') and a fifth property (A5) was added to apply the results to more non-linear models.

- (A1) λ is positive on \mathbb{R} and twice continuously differentiable.
- (A2') λ is unimodal with mode $c_\lambda^{(A2')} \in \mathbb{R}$.
- (A3') There exists a threshold $c_\lambda^{(A3')} \in \mathbb{R}$ so that the second derivative u'' of $u = \frac{1}{\lambda}$ is both injective on $(-\infty, c_\lambda^{(A3')}]$ and injective on $[c_\lambda^{(A3')}, \infty)$.
- (A4) The function $\frac{\lambda'}{\lambda}$ is non-increasing.
- (A5) $u = \frac{1}{\lambda}$ dominates z^2 asymptotically for $z \rightarrow \infty$.

If $c_\lambda^{(A2')} = c_\lambda^{(A3')}$ we will write c_λ for short. In this context condition (A2') means that there exists a $c_\lambda^{(A2')} \in \mathbb{R}$ so that λ' is positive on $(-\infty, c_\lambda^{(A2')})$ and negative on $(c_\lambda^{(A2')}, \infty)$. Hence, there is only one local maximum which is simultaneously the global maximum. So the class of intensity functions, which satisfy (A1), (A2') and (A3'), is called *class of unimodal intensity functions*. At this the condition (A3') will be needed to apply the Kiefer-Wolfowitz equivalence theorem.

The intensity functions of this class considered here have to satisfy additionally (A5). The property (A5) is

$$\lim_{z \rightarrow \infty} \left| \frac{u(z)}{z^2} \right| = \infty.$$

This means that $u(z) = \frac{1}{\lambda(z)}$ goes faster to infinity than z^2 for $z \rightarrow \infty$. The extra condition (A4) gives the log-concavity of λ . This guarantees uniqueness of the solutions in the following theorems and lemmas.

For a concise notation, we define

$$q(x_1) := \lambda(\beta_0 + \beta_1 x_1).$$

The properties (A1), (A2'), (A3'), (A4) and (A5) transfer to q for $\beta_1 > 0$, respectively, and, analogously, we set $c_q^{(\cdot)} = \frac{c_\lambda^{(\cdot)} - \beta_0}{\beta_1}$ with (\cdot) is (A2'), (A3') or empty.

It should be noted, that for fixed β^0 the following propositions do not need (A1), (A2'), (A3') and (A4) on the entire real line \mathbb{R} . It suffices to have them to hold within the ball and, in particular, on the interval $[-1, 1]$ for x_1 in the case of q and on the interval $[\beta_0 - \beta_1, \beta_0 + \beta_1]$ in the case of λ , respectively. But, for considering arbitrary β^0 , the model has to satisfy the conditions on the whole real line.

We now consider certain intensity functions: the logit model has the intensity function

$$q_{\text{logit}}(x_1) = \frac{\exp(\beta_0 + \beta_1 x_1)}{(1 + \exp(\beta_0 + \beta_1 x_1))^2}$$

and probit model has

$$q_{\text{probit}}(x_1) = \frac{\phi^2(\beta_0 + \beta_1 x_1)}{\Phi(\beta_0 + \beta_1 x_1) \cdot (1 - \Phi(\beta_0 + \beta_1 x_1))}$$

with the density function ϕ and cumulative distribution function Φ of the standard normal distribution. Both models satisfy all five conditions (A1), (A2'), (A3'), (A4), (A5) and share a common $c_\lambda^{(A2')} = c_\lambda^{(A3')} = 0$, say $c_\lambda = 0$. Analogously $c_q = -\frac{\beta_0}{\beta_1}$ for q .

Beside these two widely used models other models like the complementary log-log model, see Ford et al. (1992), with intensity function $\lambda_{\text{comp log log}}(z) = \frac{\exp(2z)}{\exp(\exp(z)) - 1}$ satisfy all five conditions — here with $c_\lambda^{(A2')} \approx 0.466011$ and $c_\lambda^{(A3')} \approx 0.049084$, but the mode $c_\lambda^{(A2')}$ and the threshold $c_\lambda^{(A3')}$ do not coincide.

We showed that if the (concise) intensity function q satisfies (A1), (A2'), (A3') and (A5) the (locally) D -optimal design $\xi^* = \xi_1^* \otimes \bar{\eta}$ is concentrated on exactly two orbits, which are the support points of the marginal design ξ_1^* . The idea of the proof is based on Biedermann et al. (2006) and Konstantinou et al. (2014).

The next theorem is the main result of the paper Radloff and Schwabe (2019b) and is reproduced (with a slight adaptation for more precision) for the readers' convenience. It characterizes the positions of the two support points of the optimal marginal design ξ_1^* .

Theorem 1 For $k \geq 2$ the simplified problem (3) with $\beta_1 > 0$ and intensity function q satisfying (A1), (A2'), (A3') and (A5) has a (locally) D -optimal marginal design ξ_1^* with exactly 2 support points x_{11}^* and x_{12}^* with $x_{11}^* > x_{12}^*$ and weights $w_1 = \xi_1^*(x_{11}^*)$ and $w_2 = \xi_1^*(x_{12}^*)$.

There are 3 cases:

- (a) If $c_q^{(A2')} > 1$ and $c_q^{(A3')} \notin [-1, 1]$, then $x_{11}^* = 1$, $w_1 = \frac{1}{k+1}$, $w_2 = \frac{k}{k+1}$ and $x_{12}^* \in (-1, 1)$ is solution of

$$\frac{q'(x_{12}^*)}{q(x_{12}^*)} = \frac{2(1 + kx_{12}^*)}{k(1 - x_{12}^{*2})}. \tag{4}$$

If additionally (A4) is satisfied, the solution x_{12}^* is unique.

- (b) If $c_q^{(A2')} < -1$ and $c_q^{(A3')} \notin [-1, 1]$, then $x_{12}^* = -1$, $w_1 = \frac{k}{k+1}$, $w_2 = \frac{1}{k+1}$ and $x_{11}^* \in (-1, 1)$ is solution of

$$\frac{q'(x_{11}^*)}{q(x_{11}^*)} = \frac{2(-1 + kx_{11}^*)}{k(1 - x_{11}^{*2})}. \tag{5}$$

If additionally (A4) is satisfied, the solution x_{11}^* is unique.

- (c) Otherwise $c_q^{(A2')} \in [-1, 1]$ or $c_q^{(A3')} \in [-1, 1]$.

Let $x, y \in \mathbb{R}$ with $x > y$ and $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ be solution of the equation system:

$$\frac{q'(x)}{q(x)} + \frac{2}{x-y} + (k-1) \frac{q'(x)(1-x^2)(\frac{1}{2}-\alpha) + q(x)(-2x)(\frac{1}{2}-\alpha)}{q(x)(1-x^2)(\frac{1}{2}-\alpha) + q(y)(1-y^2)(\frac{1}{2}+\alpha)} = 0 \tag{6}$$

$$\frac{q'(y)}{q(y)} - \frac{2}{x-y} + (k-1) \frac{q'(y)(1-y^2)(\frac{1}{2}+\alpha) + q(y)(-2y)(\frac{1}{2}+\alpha)}{q(x)(1-x^2)(\frac{1}{2}-\alpha) + q(y)(1-y^2)(\frac{1}{2}+\alpha)} = 0 \tag{7}$$

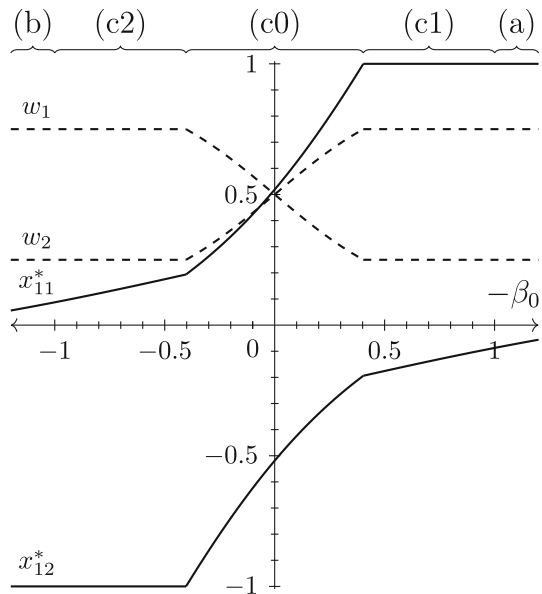
$$\frac{1}{\frac{1}{2}-\alpha} - \frac{1}{\frac{1}{2}+\alpha} + (k-1) \frac{q(x)(1-x^2) - q(y)(1-y^2)}{q(x)(1-x^2)(\frac{1}{2}-\alpha) + q(y)(1-y^2)(\frac{1}{2}+\alpha)} = 0 \tag{8}$$

- (c0) If $x, y \in (-1, 1)$ with $x > y$ and $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ is a solution of the equation system, the orbit positions are $x_{11}^* = x$, $x_{12}^* = y$ with weights $w_1 = \frac{1}{2} - \alpha$ and $w_2 = \frac{1}{2} + \alpha$.

- (c1) If $x \geq 1$ and $y \in (-1, 1)$, then $x_{11}^* = 1$, $w_1 = \frac{1}{k+1}$, $w_2 = \frac{k}{k+1}$ and $x_{12}^* \in (-1, 1)$ is the solution of the Eq. (4).

- (c2) If $y \leq -1$ and $x \in (-1, 1)$, then $x_{12}^* = -1$, $w_1 = \frac{k}{k+1}$, $w_2 = \frac{1}{k+1}$ and $x_{11}^* \in (-1, 1)$ is the solution of the Eq. (5).

Fig. 1 Logit model for $k = 3$ and $\beta_1 = 1$: dependence of x_{11}^* and x_{12}^* (solid lines) and the corresponding weights w_1 and $w_2 = 1 - w_1$ (dashed lines) on $c_q = -\beta_0 \in [-1.2, 1.2]$



Remark 1 Instead of presenting the whole theorem for $k = 1$, only the two main changes in case (c) should be mentioned. So the weights are always $w_1 = w_2 = \frac{1}{2}$ and the equation system (6)–(8) is replaced by

$$\frac{q'(x)}{q(x)} + \frac{2}{x - y} = 0 \quad \text{and} \quad \frac{q'(y)}{q(y)} - \frac{2}{x - y} = 0. \tag{9}$$

To illustrate this complex issue we revisit the logit model in dimension $k = 3$ with $\beta_1 = 1$. We (numerically) plot the orbit positions x_{11}^* and x_{12}^* and corresponding weights w_1 and $w_2 = 1 - w_1$ depending on $-\beta_0 = -\frac{\beta_0}{\beta_1} = c_q$, see Fig. 1. The cases (a) and (b) are in accordance with the results from Radloff and Schwabe (2019a) because the intensity function is monotonic within the ball. The cases (c1) and (c2) yield marginal extremum solutions which are identical to (a) and (b). So for these four cases there exists always an exact minimally supported (locally) D -optimal design. It consists of a pole point in $x_1 = -1$ or $x_1 = 1$ and the k vertices of a (regular) simplex which is maximally inscribed in the non-degenerated orbit at $x_1 = x_{11}^*$ or $x_1 = x_{12}^*$, respectively.

But the case (c0) is more problematic because the (locally) D -optimal (generalized) design consists of two non-degenerated orbits and additionally the weights are rarely appropriate for a discretization. In Radloff and Schwabe (2019b) we showed two examples for the logit model ($k = 3, \beta_1 = 1$) from which we derived (nearly) exact designs.

For $-\beta_0 = 0$ the two orbit positions are symmetrical around 0, that is $x_{11}^* = -x_{12}^* \approx 0.52$, and the weights are $\xi_1^*(x_{11}^*) = \xi_1^*(x_{12}^*) = \frac{1}{2}$. These two orbits were discretized

by two 2-dimensional simplices—overall 6 equally weighted support points, see Fig. 2 (left image).

For $-\beta_0 = -0.1$ the solutions are $x_{11}^* \approx 0.42, x_{12}^* \approx -0.62$ and $\xi_1^*(x_{11}^*) \approx 0.4297$, while $0.4297 \approx \frac{3}{7}$. We chose the rounded design ξ^{\approx} with the same support points x_{11}^* and x_{12}^* but with the marginal design $\xi_1^{\approx}(x_{11}^*) = \frac{3}{7}$ and $\xi_1^{\approx}(x_{12}^*) = \frac{4}{7}$. So it was possible to substitute one orbit by the vertices of a 2-dimensional simplex (3 points—an equilateral triangle) and one by the vertices of a 2-dimensional cube or cross polytope (4 points—a square). Because of rounding the design ξ^{\approx} is not optimal but exact and has a high D -efficiency, which compares the rounded design ξ^{\approx} and the optimal design $\xi_{\beta^0}^*$ with respect to β^0 —here $p = k + 1 = 4$ and $\beta^0 = (0.1, 1, 0, 0)^T$:

$$\text{Eff}_D(\xi^{\approx}, \beta^0) = \left(\frac{\det(\mathbf{M}(\xi^{\approx}, \beta^0))}{\det(\mathbf{M}(\xi_{\beta^0}^*, \beta^0))} \right)^{\frac{1}{p}} \approx 0.999757.$$

These designs have the following properties, which are unsatisfactory. On the one hand the number of support points is not minimal. On the other hand only special cases have appropriate rational weights which allow a discretization or otherwise the optimality is lost by rounding. Therefore we want to construct minimal supported exact designs for the case (c0) in this paper which will be (highly) efficient if not optimal.

But we start with the reduction of the system of three equations in Theorem 1 to only one single equation for special unimodal intensity functions—symmetrical unimodal intensity functions. They occur, for example, in binary response models with logit and probit link.

4 Optimal design for symmetrical unimodal intensity functions

An interesting observation was made in the discussion section in Radloff and Schwabe (2019b). For models with unimodal intensity function in which the mode and the threshold coincide ($c_{\lambda}^{(A2')} = c_{\lambda}^{(A3')} = c_{\lambda}$) and which are symmetrical, also the two orbit positions are symmetrical in a certain way, which we want to investigate here. For one dimension this has been considered and shown in Ford et al. (1992, Sects. 6.5 and 6.6), but this proof cannot be extended to higher dimensions directly.

A unimodal intensity function in which the mode and the threshold coincide ($c_{\lambda}^{(A2')} = c_{\lambda}^{(A3')} = c_{\lambda}$) will be called *symmetrical* to c_{λ} if

$$\lambda(c_{\lambda} + z) = \lambda(c_{\lambda} - z)$$

for all $z \in \mathbb{R}$.

The intensity functions of the logit and probit models are symmetrical with $c_{\lambda} = 0$. But the unimodal intensity function of the complementary log-log model has $c_{\lambda}^{(A2')} \neq c_{\lambda}^{(A3')}$ and cannot be symmetrical for this reason.

In the present paper we focus only on the logit and probit models as practically important examples and most commonly used models with symmetrical unimodal intensity function. But it is conceivable that there are more models of this type, particularly with regarding to binary response. Assuming Y_i as Bernoulli distributed with success probability $p_i = F(\mathbf{f}'(\mathbf{x}_i)^\top \boldsymbol{\beta})$, where F is a strictly increasing, continuously differentiable cumulative distribution function and $f = F'$ is the corresponding density function, the intensity function is

$$\lambda(z) = \frac{f^2(z)}{F(z)(1 - F(z))}.$$

Then symmetry will be inherited: if the density function is symmetrical, the intensity function will be symmetrical, too.

If the density function has additionally a local extremum at the symmetry line, the intensity function will also have a local extremum there. It has to be checked separately whether this is the only (local) maximum.

Lemma 2 *Let the intensity function λ be symmetrical to c_λ in the situation of Theorem 1 (c0).*

- For $\beta_0 \neq c_\lambda$ let r solve

$$\frac{\lambda'(c_\lambda + r)}{\lambda(c_\lambda + r)} = - \frac{A(k, r, c, \beta_1)}{(k+1)r(r+c-\beta_1)(r+c+\beta_1)(r-c+\beta_1)(r-c-\beta_1)} \tag{10}$$

with

$$A(k, r, c, \beta_1) := -2kr^2(\beta_1^2 + c^2 - r^2) + (\beta_1^2 - c^2 - r^2)^2 - 4c^2r^2 + (\beta_1^2 - c^2 + r^2)\sqrt{(\beta_1^2 - c^2 - r^2)^2 + 4(k^2 - 1)c^2r^2}$$

and $c := c_\lambda - \beta_0$. Then

$$x = \frac{c}{\beta_1} + \frac{r}{\beta_1}, \tag{11}$$

$$y = \frac{c}{\beta_1} - \frac{r}{\beta_1}, \tag{12}$$

$$\alpha = \frac{-(\beta_1^2 - c^2 - r^2) + \sqrt{(\beta_1^2 - c^2 - r^2)^2 + 4(k^2 - 1)c^2r^2}}{4(k+1)cr} \tag{13}$$

is a solution of the equation system (6)–(8).

- For $\beta_0 = c_\lambda$ a solution of (6)–(8) is $x = \frac{r}{\beta_1}$, $y = -\frac{r}{\beta_1}$ and $\alpha = 0$, where r is the solution of

$$\frac{\lambda'(c_\lambda + r)}{\lambda(c_\lambda + r)} = - \frac{2(\beta_1^2 - kr^2)}{(k+1)r(\beta_1^2 - r^2)}. \tag{14}$$

Remark 2 For the particular case $k = 1$, cf. Remark 1, let λ be symmetrical to c_λ . Then $x = \frac{c_\lambda - \beta_0}{\beta_1} + \frac{r}{\beta_1}$ and $y = \frac{c_\lambda - \beta_0}{\beta_1} - \frac{r}{\beta_1}$ with r solution of

$$\frac{\lambda'(c_\lambda + r)}{\lambda(c_\lambda + r)} = -\frac{1}{r} \tag{15}$$

solve the equation system (9).

Lemma 2, whose proof sketch can be found in Appendix B, and Remark 2 in combination with Theorem 1 give (locally) D -optimal designs for models with symmetrical unimodal intensity functions. As a result we reduced the system of Eqs. (6)–(8) to only one single Eq. (10).

But the question remains whether condition (A4) can guarantee a unique solution as in Theorem 1(a) and (b) because Theorem 1(c), especially (c0), tells nothing about the uniqueness of the positions of the two orbits. Without uniqueness there may be more than one optimal design of this shape. Before dealing with that, we want to add a remark on the range of values for r in Lemma 2, so that there are two non-degenerated orbits.

Remark 3 If the system of Eqs. (6)–(8) in Theorem 1(c0) has a solution with two inner support points for the marginal design, it is required that $x, y \in (-1, 1)$ and, hence,

$$-1 < \frac{c_\lambda - \beta_0}{\beta_1} \pm \frac{r}{\beta_1} < 1$$

must be valid. This leads with $\beta_1 > 0$ to $r \in (-(c_\lambda - \beta_0) - \beta_1, -(c_\lambda - \beta_0) + \beta_1)$ and $r \in ((c_\lambda - \beta_0) - \beta_1, (c_\lambda - \beta_0) + \beta_1)$. Consequently, both intervals must overlap. This happens for $c_\lambda - \beta_0 > 0$ at $0 < c_\lambda - \beta_0 < \beta_1$ and for $c_\lambda - \beta_0 < 0$ at $-\beta_1 < c_\lambda - \beta_0 < 0$. Thus $c_\lambda - \beta_0 \in (-\beta_1, \beta_1)$ and in particular $\beta_1^2 > (c_\lambda - \beta_0)^2$ must hold. Then r is in the interval $(|c_\lambda - \beta_0| - \beta_1, -|c_\lambda - \beta_0| + \beta_1)$. But Theorem 1(c) needs $x > y$ and consequently $r > 0$. Hence, $r \in (0, -|c_\lambda - \beta_0| + \beta_1)$.

This remains valid in particular for $\beta_0 = c_\lambda$, i.e. $c_\lambda - \beta_0 = 0$. So $r \in (-\beta_1, \beta_1)$. With $r > 0$ it is $r \in (0, \beta_1)$.

Lemma 3 In situation of Lemma 2 let the intensity function λ additionally satisfy condition (A4), then Eq. (10), whose right hand side is continuously continued in $-|c_\lambda - \beta_0| + \beta_1$, has a unique solution in $r \in (0, |c_\lambda - \beta_0| + \beta_1)$.

This also holds for $\beta_0 = c_\lambda$ and Eq. (14), which has exactly one solution in $r \in (0, \beta_1)$.

Remark 4 For $k = 1$, cf. Remark 2, and for an intensity function satisfying (A4) there is only one solution of (15).

The proof of Lemma 3 is sketched in Appendix B. Lemma 3 guarantees a unique solution in $r \in (0, |c_\lambda - \beta_0| + \beta_1)$. But Remark 3 points out that for Theorem 1 (c0) we need $r \in (0, -|c_\lambda - \beta_0| + \beta_1)$. This means that the unique solution can result in the two-orbit case or in the one-orbit one-pole case of Theorem 1 (c).

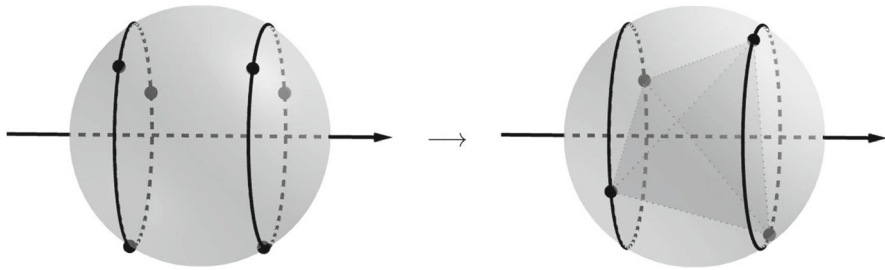


Fig. 2 Logit model for $k = 3$, $\beta_1 = 1$ and $-\beta_0 = 0$: discretized (locally) D -optimal designs with 6 or 4 support points

5 Minimally supported designs

In the situation of Theorem 1(a), (b), (c1) and (c2) the designs have always the minimal number of support points to estimate the parameter vector β . These are $k + 1$ support points.

In Radloff and Schwabe (2019b) revisited here in the introductory section we indicated exemplarily a (locally) D -optimal design for the logit model on the 3-dimensional ball with $-\beta_0 = 0$ and $\beta_1 = 1$. This design consists of six support points which are the vertices of two regular 2-dimensional simplices—equilateral triangles, see Fig. 2 (left image). But this is not the minimum of support points to estimate the four parameters.

So the question arises whether it is possible to reduce the number of support points as it can be found in the concept of *fractional factorial designs*, see e. g. Pukelsheim (1993, Sect. 15.11). Instead of using all vertices of the hypercube $[-1, 1]^k$ as in the full factorial design the fractional factorial design picks only a special percentage of these points. For $k = 3$

$$\left((-1, -1, 1)^\top, (-1, 1, -1)^\top, (1, -1, -1)^\top, (1, 1, 1)^\top \right)$$

represents a 2^{3-1} -fractional factorial design.

Here, we do not want to pick four of the six points, but we want to use the orthogonality of the spaces spanned by the points (without the x_1 -component) in the two orbits ($x_1 = -1$ and $x_1 = 1$) of the given 2^{3-1} -fractional factorial design. Here $\text{span}\{(-1, 1)^\top, (1, -1)^\top\} \perp \text{span}\{(-1, -1)^\top, (1, 1)^\top\}$. The idea is illustrated in Fig. 2 (right image). The spanned spaces by points (without the x_1 -component) in the orbits are orthogonal to each other. And all points span a simplex.

As stated above a (generalized) design ξ which is rotationally invariant with fixed x_1 (this means it is invariant with respect to all orthogonal transformations in the orthogonal group $O(k)$ which do not change the x_1 -component) and which has all mass on the unit sphere can be decomposed into a marginal design ξ_1 on $[-1, 1]$ and a probability kernel $\bar{\eta}$ (conditional design), i. e. $\xi = \xi_1 \otimes \bar{\eta}$. For fixed x_1 the kernel $\bar{\eta}(x_1, \cdot)$ is the uniform distribution on the surface of a $(k - 1)$ -dimensional ball with radius $\sqrt{1 - x_1^2}$ —the radius of the orbit at position x_1 . If $x_1 \in \{-1, 1\}$, the $(k - 1)$ -

dimensional ball with the uniform distribution reduces to a single point and represents only a one-point measure. Remembering $q(x_1) = \lambda(\beta_0 + \beta_1 x_1)$ the related information matrix, see Radloff and Schwabe (2019a), is

$$M(\xi_1 \otimes \bar{\eta}, \beta^0) = \left(\begin{array}{c|c} \begin{matrix} \int q \, d\xi_1 & \int q \, \text{id} \, d\xi_1 \\ \int q \, \text{id} \, d\xi_1 & \int q \, \text{id}^2 \, d\xi_1 \end{matrix} & \mathbb{O}_{2 \times (k-1)} \\ \hline \mathbb{O}_{(k-1) \times 2} & \frac{1}{k-1} \int q (1 - \text{id}^2) \, d\xi_1 \, \mathbb{I}_{k-1} \end{array} \right) \tag{16}$$

with the identity function $\text{id} (\text{id}(x_1) = x_1)$ and the parameter vector $\beta^0 = (\beta_0, \beta_1, 0, \dots, 0)^\top$.

The information matrix for a design on the k -dimensional unit sphere \mathbb{S}_{k-1} , which is based on exactly two orbits, can be determined analogously to this result. Additionally the uniform distribution does not cover the the full orbits but only sub-spheres.

Lemma 4 *Let ξ_1 be the two-point measure in x_{11} and x_{12} with $\xi_1(x_{11}) = \frac{1}{2} - \alpha$ and $\xi_1(x_{12}) = \frac{1}{2} + \alpha$ with $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. Further let $\bar{\eta}(x_{11}, \cdot)$ be a uniform distribution on $\mathbb{S}_{m-2}(\sqrt{1 - x_{12}^2}) \times \{0\}^{k-m}$ and likewise $\bar{\eta}(x_{12}, \cdot)$ be a uniform distribution on $\{0\}^{m-1} \times \mathbb{S}_{k-m-1}(\sqrt{1 - x_{12}^2})$. Then the information matrix is*

$$M(\xi_1 \otimes \bar{\eta}, \beta^0) = \left(\begin{array}{c|c} \begin{matrix} \int q \, d\xi_1 & \int q \, \text{id} \, d\xi_1 \\ \int q \, \text{id} \, d\xi_1 & \int q \, \text{id}^2 \, d\xi_1 \end{matrix} & \mathbb{O}_{2 \times (k-1)} \\ \hline \mathbb{O}_{(k-1) \times 2} & \begin{matrix} c_1 \mathbb{I}_{m-1} & \mathbb{O}_{(m-1) \times (k-m)} \\ \mathbb{O}_{(k-m) \times (m-1)} & c_2 \mathbb{I}_{k-m} \end{matrix} \end{array} \right) \tag{17}$$

with $c_1 = \frac{1}{m-1} q(x_{11}) (1 - x_{11}^2) (\frac{1}{2} - \alpha)$ and $c_2 = \frac{1}{k-m} q(x_{12}) (1 - x_{12}^2) (\frac{1}{2} + \alpha)$.

Now the optimality case in Theorem 1 (c0) on two orbits should be used to investigate when both information matrices (16) and (17) are identical. With that both related (generalized) designs would be (locally) D -optimal.

Lemma 5 *Both information matrices (16) and (17) are identical in the situation of Theorem 1(c0) if and only if $\alpha = \frac{1}{2} - \frac{m}{k+1}$.*

The proof can be found in Appendix B.

Consequently both orbits need the weights $\xi_1(x_{11}) = \frac{m}{k+1}$ and $\xi_1(x_{12}) = \frac{k-m+1}{k+1}$ to coincide both information matrices. This allows an experimental design, which has the same value for the D -optimality criterion, consisting of two orbits with m and with $k - m + 1$ support points. This can be done by two regular simplices— one simplex in dimension $m - 1$ and one in dimension $k - m$. So the simplices are the discretizations of the uniform distributions on $\mathbb{S}_{m-2}(\sqrt{1 - x_{11}^2}) \times \{0\}^{k-m}$ and on $\{0\}^{m-1} \times \mathbb{S}_{k-m-1}(\sqrt{1 - x_{12}^2})$.

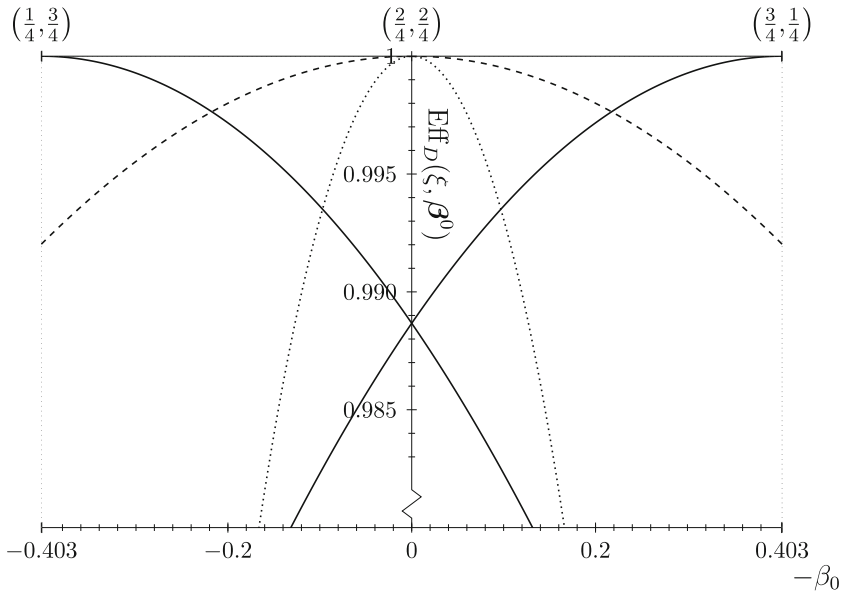


Fig. 3 Logit model with $k = 3$ and $\beta_1 = 1$: D -efficiency for type-1-designs (solid lines), type-2-designs (dotted lines) and type-3-designs (dashed lines) with exactly $k + 1 = 4$ equally weighted support points for $-\beta_0 \in (-0.403, 0.403)$

Let $S_m \in \mathbb{R}^{m \times (m+1)}$ be a matrix, where the columns represent the $m + 1$ vertices of an m -dimensional regular simplex (in \mathbb{R}^m). Then the columns of the matrix

$$\left(\begin{array}{c|c} x_{11} \mathbb{1}_m^\top & x_{12} \mathbb{1}_{k-m+1}^\top \\ \mathbf{R}_1 S_{m-1} & \mathbb{O}_{(m-1) \times (k-m+1)} \\ \hline \mathbb{O}_{(k-m) \times m} & \mathbf{R}_2 S_{k-m} \end{array} \right)$$

with arbitrary orthogonal transformations $\mathbf{R}_1 \in O(m - 1)$ and $\mathbf{R}_2 \in O(k - m)$ represent the support points of such a minimal supported design.

$$\left(\sqrt{\frac{m+1}{m}} \mathbb{1}_m + \frac{1 - \sqrt{m+1}}{m\sqrt{m}} \mathbb{1}_m \mathbb{1}_m^\top \mid -\frac{1}{\sqrt{m}} \mathbb{1}_m \right) \in \mathbb{R}^{m \times (m+1)}$$

is an example for S_m . In this notation $\mathbb{1}_m$ stands for the standard simplex which needs to be scaled and shifted appropriately so that it is in combination with the last vertex $-\frac{1}{\sqrt{m}} \mathbb{1}_m$ (last column) a regular simplex on the unit sphere \mathbb{S}_{m-1} .

Finally, we want to look at the D -efficiency, here with $\beta^0 = (\beta_0, \beta_1, 0, \dots, 0)^\top$,

$$\text{Eff}_D(\xi, \beta^0) = \left(\frac{\det(\mathbf{M}(\xi, \beta^0))}{\det(\mathbf{M}(\xi_{\beta^0}^*, \beta^0))} \right)^{\frac{1}{p}} \in [0, 1]$$

for designs ξ with exactly $p = k + 1$ equally weighted support points in the region where two non-degenerated orbits occur.

As an example, the logit model with $\beta_1 = 1$ is used to determine the D -efficiency in dimensions $k = 3$ and $k = 6$. In Figs. 3 and 4 only the regions for $-\beta_0$ with two non-degenerated orbits in the optimal design (case (c0) in Theorem 1), i.e. $-\beta_0 \in (-0.403, 0.403)$ for $k = 3$ and $-\beta_0 \in (-0.480, 0.480)$ for $k = 6$, are plotted.

For this purpose, three different types of exact designs are compared with the (locally) D -optimal design $\xi_{\beta^0}^*$. The optimal design is a generalized design with real-valued weights. Therefore it cannot be discretized as an exact design in general.

First, the two optimal exact designs with one pole and one orbit, which are discretized as a regular $(k - 1)$ -dimensional simplex, are used for comparison. The orbit position remains unchanged and is determined at the boundary values $-\beta_0 \approx \pm 0.403$ or $-\beta_0 \approx \pm 0.480$ for $k = 3$ or $k = 6$, respectively. See the solid lines for these *type-1-designs* in both figures.

Second, the designs with the same orbit position as the associated design which is (locally) optimal for $-\beta_0$ are the next alternative. Only the weights were rounded/shifted to integral multiples of $\frac{1}{k+1}$. See the dotted lines for these *type-2-designs*.

Third, the designs with fixed design weights which are integral multiples of $\frac{1}{k+1}$ represent the last model category. So only the positions of the orbits have to be optimized with these fixed design weights. This can be done by solving only the Eqs. (6) and (7) with the selected weights in Theorem 1(c). Equation (8) is omitted. See the dashed lines for these *type-3-designs* in both plots.

The Fig. 3 reveals for dimension $k = 3$ that there are only three positions in the range $-\beta_0 \in (-0.403, 0.403)$ where (locally) D -optimal designs with the minimal number of support points, which are four points, exist. For $-\beta_0 \approx -0.403$ this is the design (type-1-design) consisting of the pole $x_{12}^* = -1$ and one orbit at x_{11}^* with three points as vertices of an equilateral triangle. Then for $-\beta_0 = 0$ there are two orbits with two points each. And, at $-\beta_0 \approx 0.403$ the design (type-1-design) consists of one orbit at x_{12}^* with three equally weighted support points and the pole $x_{11}^* = 1$. In the span between these optimality positions the considered discretizations provide a fairly high efficiency. Using the transition directly from pole and orbit to orbit and pole, the efficiency is always greater than 0.988 (intersection of the solid lines, both type-1-designs). If the two orbits are also discretized in between, the efficiency is greater than 0.993 (intersection of dotted line and solid lines, type-2- and type-1-designs) or even greater than 0.997 (intersection of dashed line and solid lines, type-3- and type-1-designs).

For dimension $k = 6$, see Fig. 4, an efficiency of more than 0.986 is possible by stepping directly from pole and orbit with six support points to orbit with six design points and pole (both type-1-designs). If the intermediate steps (two orbits with 2 and 5 points, 3 and 4 points, 4 and 3 points as well as 5 and 2 points) are used, then by simple rounding of the weights to integral multiples of $\frac{1}{k+1}$ an efficiency greater than 0.995 (dotted lines and solid lines, type-2- and type-1-designs) and with additional optimization of the orbit positions even greater than 0.999 (dashed lines and solid lines, type-3- and type-1-designs) can be achieved.

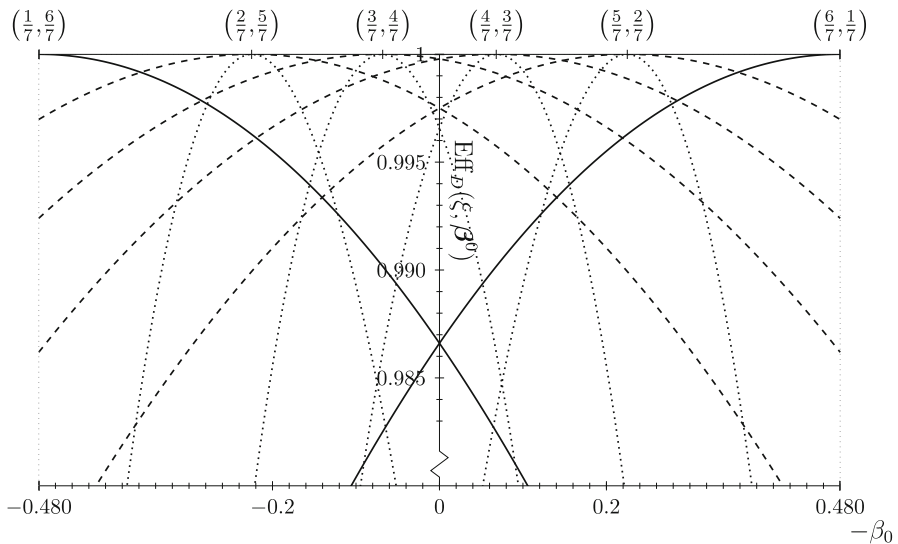


Fig. 4 Logit model with $k = 6$ and $\beta_1 = 1$: D -efficiency for type-1-designs (solid lines), type-2-designs (dotted lines) and type-3-designs (dashed lines) with exactly $k + 1 = 7$ equally weighted support points for $-\beta_0 \in (-0.480, 0.480)$

6 Conclusion

In summary it can be postulated that very efficient designs are generated based on only $k + 1$ design points which is the minimal number of support points to estimate the parameter vector. It seems that higher dimensions enable designs with higher D -efficiency, in particular using the third option of discretization. Here we only considered designs with exactly two orbits. Thus it cannot be excluded that there are designs with a better efficiency or even (locally) optimal designs which are supported by exactly $k + 1$ points. Maybe these designs have support points which lie not on the orbit but are jittered a little bit. This as well as a potential lower efficiency bound needs further investigations.

On the other side the reduction of the equation system to one single equation for determining (locally) D -optimal design for symmetrical unimodal intensity functions is a nice feature and can help to decrease computing costs.

Also the question of optimal designs on the ball with respect to other optimality criteria should be considered in future.

Finally, we want to emphasize that the established designs do not only work for the unit ball. By using the concept of equivariance for linear transformations, say scaling, reflecting and rotating, the class of design regions can be extended to k -dimensional balls with arbitrary radius or any k -dimensional ellipsoid.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Notation

\mathbb{B}_k	k -dimensional unit ball
$\mathbb{B}_k(r)$	k -dimensional ball with radius r
\mathbb{S}_{k-1}	unit sphere, which is the surface of \mathbb{B}_k
$\mathbb{S}_{k-1}(r)$	sphere with radius r , which is the surface of $\mathbb{B}_k(r)$
$\mathbb{0}_k$	k -dimensional zero-vector
$\mathbb{0}_{k_1 \times k_2}$	$(k_1 \times k_2)$ -dimensional zero-matrix
$\mathbb{1}_k$	k -dimensional one-vector
\mathbb{I}_k	$(k \times k)$ -dimensional identity matrix
id	identity function, $\text{id}(x) = x$
$O(k)$	The orthogonal group in \mathbb{R}^k , which is the set of all orthogonal linear transformations and can be represented by $(k \times k)$ -matrices. So $O(k) = \{A \in \mathbb{R}^{k \times k} : A^\top A = \mathbb{I}_k\}$ such that $\det(A) \in \{-1, 1\}$.

Appendix B: Proofs

Sketch of the Proof of Lemma 2 By plugging (11) and (12) into (8) and using the symmetry to simplify, we get

$$\frac{-2\alpha(4cr\alpha + (\beta_1^2 - c^2 - r^2)) + 4(k-1)cr(\frac{1}{2} - \alpha)(\frac{1}{2} + \alpha)}{(\frac{1}{2} - \alpha)(\frac{1}{2} + \alpha)(4cr\alpha + (\beta_1^2 - c^2 - r^2))} = 0.$$

In the numerator there is a polynomial of degree two in α with the two roots $\alpha_{\mp}(r)$ depending on r :

$$\alpha_{\mp}(r) := \frac{-(\beta_1^2 - c^2 - r^2) \mp \sqrt{(\beta_1^2 - c^2 - r^2)^2 + 4(k+1)(k-1)c^2r^2}}{4(k+1)cr}.$$

Now we examine the values of $\alpha_{\mp}(r)$ depending on r . Only $-|c| - \beta_1, |c| - \beta_1, -|c| + \beta_1$ or $|c| + \beta_1$ can solve the expression $\alpha_{\mp}(r) = \pm \frac{1}{2}$. But $-|c| - \beta_1$ and $|c| + \beta_1$ are not

in the interesting region for r . We have

$$\alpha_{-}(\pm(|c| - \beta_1)) = \pm \frac{1}{2} \operatorname{sign}(c) \quad \text{and} \quad \alpha_{+}(\pm(|c| - \beta_1)) = \mp \frac{1}{2} \operatorname{sign}(c) \frac{k - 1}{k + 1}.$$

Because of $\lim_{r \nearrow 0} \alpha_{-}(r) = \operatorname{sign}(c)\infty$ and $\lim_{r \searrow 0} \alpha_{-}(r) = -\operatorname{sign}(c)\infty$ the root $\alpha_{-}(r)$ has in the interval $r \in [|c| - \beta_1, -|c| + \beta_1]$ only values outside $(-\frac{1}{2}, \frac{1}{2})$. Hence, $\alpha_{-}(r)$ is not a relevant root.

Since $\lim_{r \rightarrow 0} \alpha_{+}(r) = 0$ the discontinuity of the root $\alpha_{+}(r)$ in $r = 0$ can be removed. So $\alpha_{+}(r)$ has only values in $(-\frac{1}{2}, \frac{1}{2})$ on the interval $r \in [|c| - \beta_1, -|c| + \beta_1]$ and $\alpha_{+}(r)$, which is (13), is the only relevant root.

After inserting (11) and (12) into (6) as well as (11) and (12) into (7) and subtracting both obtained equations and simplifying by using the symmetry, we get

$$\frac{(k + 1) \lambda'(c_{\lambda} + r)}{\lambda(c_{\lambda} + r)} = -(k - 1) \frac{-2r + \alpha \cdot 4c}{(\beta_1^2 - c^2 - r^2) + \alpha \cdot 4cr} - \frac{2}{r}.$$

Equation (10) follows by plugging $\alpha_{+}(r)$ as α into it and by some simplifications.

For $\beta_0 = c_{\lambda}$, i.e. $c = c_{\lambda} - \beta_0 = 0$, we get directly $\alpha = 0$ by inserting $x = \frac{r}{\beta_1}$ and $y = -\frac{r}{\beta_1}$ in (8) and exploiting the symmetry. This is inserted in (6) and in (7). The difference between these two equations results in (14).

Sketch of the Proof of Lemma 3 This proof is a lot of curve sketching. We start with $\beta_0 \neq c_{\lambda}$. The denominator of the right hand side of (10) has five roots in r . $-|c_{\lambda} - \beta_0| - \beta_1 < 0$ and $|c_{\lambda} - \beta_0| - \beta_1 < 0$ are not in the considered interval $(0, |c_{\lambda} - \beta_0| + \beta_1)$. In $r = -|c_{\lambda} - \beta_0| + \beta_1$ there is a discontinuity which can be removed. In $r = 0$ and in $r = |c_{\lambda} - \beta_0| + \beta_1$ there are two poles. Analyzing these poles for the considered interval we see that the values start from $-\infty$ ($r \searrow 0$) and go up to $+\infty$ ($r \nearrow |c_{\lambda} - \beta_0| + \beta_1$). Sophisticated curve sketching shows that the right hand side of (10) is strictly monotonically increasing on $(0, |c_{\lambda} - \beta_0| + \beta_1)$. So it is strictly monotonically increasing and covers $(-\infty, \infty)$. In combination with (A4) for the left hand side of (10) (monotonically decreasing) there is exactly one solution.

For $\beta_0 = c_{\lambda}$ we can mention that the right hand side of (14) is also strictly monotonically increasing on $(0, \beta_1)$. Hence, there is only one solution.

An analogue result holds for the situation in Remark 4.

Proof of Lemma 5 Rearranging Eq. (8) equivalently in two ways gives

$$q(x_{12})(1 - x_{12}^2) \left(\frac{1}{2} + \alpha\right) = q(x_{11})(1 - x_{11}^2) \left(\frac{1}{2} - \alpha\right) \frac{k \left(\frac{1}{2} + \alpha\right) - \left(\frac{1}{2} - \alpha\right)}{k \left(\frac{1}{2} - \alpha\right) - \left(\frac{1}{2} + \alpha\right)} \quad \text{and}$$

$$q(x_{11})(1 - x_{11}^2) \left(\frac{1}{2} - \alpha\right) = q(x_{12})(1 - x_{12}^2) \left(\frac{1}{2} + \alpha\right) \frac{k \left(\frac{1}{2} - \alpha\right) - \left(\frac{1}{2} + \alpha\right)}{k \left(\frac{1}{2} + \alpha\right) - \left(\frac{1}{2} - \alpha\right)}.$$

The two denominators are zero if and only if $\alpha = \frac{1}{2} - \frac{1}{k+1}$ and $\alpha = \frac{1}{2} - \frac{k}{k+1}$, respectively. But this cannot happen to non-degenerated orbits because $\frac{1}{2} - \frac{k}{k+1} < \alpha < \frac{1}{2} - \frac{1}{k+1}$.

Putting both equations into the diagonal entry of the information matrix (16) yield

$$\begin{aligned} & \frac{1}{k-1} \int q(1 - \text{id}^2) d\xi_1 \\ &= q(x_{11})(1 - x_{11}^2) \left(\frac{1}{2} - \alpha\right) \left[\frac{1}{k-1} + \frac{1}{k-1} \cdot \frac{k(\frac{1}{2} + \alpha) - (\frac{1}{2} - \alpha)}{k(\frac{1}{2} - \alpha) - (\frac{1}{2} + \alpha)} \right] \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{k-1} \int q(1 - \text{id}^2) d\xi_1 \\ &= q(x_{12})(1 - x_{12}^2) \left(\frac{1}{2} - \alpha\right) \left[\frac{1}{k-1} \cdot \frac{k(\frac{1}{2} - \alpha) - (\frac{1}{2} + \alpha)}{k(\frac{1}{2} + \alpha) - (\frac{1}{2} - \alpha)} + \frac{1}{k-1} \right] \end{aligned}$$

They are identical to the diagonal entries of the information matrix (17) in Lemma 4 if and only if

$$\begin{aligned} & \frac{1}{k-1} + \frac{1}{k-1} \cdot \frac{k(\frac{1}{2} + \alpha) - (\frac{1}{2} - \alpha)}{k(\frac{1}{2} - \alpha) - (\frac{1}{2} + \alpha)} = \frac{1}{m-1} \text{ and} \\ & \frac{1}{k-1} \cdot \frac{k(\frac{1}{2} - \alpha) - (\frac{1}{2} + \alpha)}{k(\frac{1}{2} + \alpha) - (\frac{1}{2} - \alpha)} + \frac{1}{k-1} = \frac{1}{k-m} \end{aligned}$$

which are both equivalent to $\alpha = \frac{1}{2} - \frac{m}{k+1}$.

References

- Biedermann S, Dette H, Zhu W (2006) Optimal designs for dose-response models with restricted design regions. *J Am Stat Assoc* 101:747–759
- Box GEP, Draper NR (1987) Empirical model building and response surfaces. Wiley, New York
- Dette H, Melas VB, Pepelyshev A (2005) Optimal designs for three-dimensional shape analysis with spherical harmonic descriptors. *Ann Stat* 33:2758–2788
- Dette H, Melas VB, Pepelyshev A (2007) Optimal designs for statistical analysis with zernike polynomials. *Statistics* 41:453–470
- Farrell RH, Kiefer J, Walbran A (1967) Optimum multivariate designs. In: Proceedings of the fifth berkeley symposium on mathematical statistics and probability, volume 1: statistics. University of California Press, Berkeley, pp 113–138
- Ford I, Torsney B, Wu C (1992) The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J Roy Stat Soc B* 54:569–583
- Hirao M, Sawa M, Jimbo M (2015) Constructions of ϕ_p -optimal rotatable designs on the ball. *Sankhya A* 77:211–236
- Kiefer JC (1959) Optimum experimental designs. *J R Stat Soc B* 21:272–304
- Kiefer JC (1961) Optimum experimental designs V, with applications to systematic and rotatable designs. In: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability. Univ of California Press, pp 381–405
- Konstantinou M, Biedermann S, Kimber A (2014) Optimal designs for two-parameter nonlinear models with application to survival models. *Stat Sin* 24:415–428
- Lau TS (1988) D -optimal designs on the unit q -ball. *J Stat Plann inference* 19:299–315
- Pukelsheim F (1993) Optimal design of experiments. Wiley, New York

- Radloff M, Schwabe R (2016) Invariance and equivariance in experimental design for nonlinear models. In: Kunert J, Müller CH, Atkinson AC (eds) mODa 11. Springer, Berlin, pp 217–224
- Radloff M, Schwabe R (2019a) Locally D -optimal designs for non-linear models on the k -dimensional ball. *J Stat Plan Inference* 203:106–116
- Radloff M, Schwabe R (2019b) Locally D -optimal designs for a wider class of non-linear models on the k -dimensional ball with applications to logit and probit models. *Stat Pap* 60:165–177
- Schmidt D, Schwabe R (2017) Optimal design for multiple regression with information driven by the linear predictor. *Stat Sin* 27:1371–1384

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.