Statistica Sinica Preprint No: SS-2015-0385R2	
Title	Optimal Design for Multiple Regression with Information Driven by the Linear Predictor
Manuscript ID	SS-2015-0385R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202015.0385
Complete List of Authors	Dennis Schmidt and Rainer Schwabe
Corresponding Author	Dennis Schmidt
E-mail	dennis.schmidt@ovgu.de

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

Statistica Sinica

Optimal Design for Multiple Regression with Information Driven by the Linear Predictor

Dennis Schmidt and Rainer Schwabe

Otto-von-Guericke-Universität Magdeburg

Abstract: In this paper we consider nonlinear models with an arbitrary number of covariates for which the information additionally depends on the value of the linear predictor. We establish the general result that for many optimality criteria the support points of an optimal design lie on the edges of the design region, if this design region is a polyhedron. Based on this result we show that under certain conditions the D-optimal designs can be constructed from the D-optimal designs in the marginal models with single covariates. This can be applied to a broad class of models, which include the Poisson, the negative binomial as well as the proportional hazards model with both type I and random censoring.

Key words and phrases: multiple regression model, D-optimality, censored data, proportional hazards model, generalized linear models.

1. Introduction

In this paper we determine locally D-optimal designs for a large class of models with an arbitrary number of covariates. Since the models considered are

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

nonlinear, the optimal designs depend on the unknown parameters. In accordance with Chernoff (1953) such designs, which are optimal for a prespecified parameter value, are called locally optimal. We show that the D-optimal designs can be constructed from the D-optimal designs in the marginal models with single covariates. In Section 4 we apply these results to a broad class of models, for which the intensity function of the information matrix depends on the value of the linear predictor. Such a model with only one covariate was considered by Konstantinou et al. (2014), who computed the D-optimal design and the c-optimal design for the effect parameter when the design region is an interval, and by Schmidt and Schwabe (2015), who extended these results to a discrete design region. Our conditions on the intensity function are satisfied by such models as the Poisson and the negative binomial. For the Poisson model Russell et al. (2009) determined D-optimal designs, where the number of covariates is arbitrary. This result follows from ours as a special case. For the negative binomial model Rodríguez-Torreblanca and Rodríguez-Díaz (2007) determined D- and c-optimal designs for a single covariate. In the context of measuring human intelligence Graßhoff et al. (2016) considered the Poisson-Gamma model, which is equivalent to the negative binomial model, determining D-optimal designs with two binary covariates. We consider this model with two continuous covariates and show that the D-optimal design is much better than the natural design with equal weights

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

on all four possible design points. Another model covered by our model is the proportional hazards model with type I and random censoring (cf. Konstantinou et al. (2014)), which will be discussed in Section 5.

2. Model specifications

The information matrix depends on the control variables, which can be chosen by the experimenter. Since under mild regularity conditions the inverse of the Fisher information matrix is proportional to the asymptotic covariance of the asymptotically efficient maximum likelihood estimator, we want to maximize the information matrix in some sense in order to find the optimal choice of control variables for obtaining the most precise parameter estimates. We determine approximate designs (cf. Silvey (1980, p. 15))

$$\xi = \begin{Bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_m \\ \omega_1 & \omega_2 & \dots & \omega_m \end{Bmatrix},$$

where $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ are distinct values of the control variables from a given design region $\mathscr{X}$ and $\omega_1, \dots, \omega_m$ are the corresponding weights satisfying $0 \le \omega_i \le 1$ for $i = 1, \dots, m$ and $\sum_{i=1}^m \omega_i = 1$. An approximate design can be represented by a probability measure on $\mathscr{X}$ with finite support. The information matrix $\boldsymbol{M}(\xi, \boldsymbol{\beta})$ of a design ξ is (cf. Silvey (1980, p. 53))

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \int_{\mathscr{X}} \boldsymbol{I}(\boldsymbol{x}, \boldsymbol{\beta})\, \xi(\mathrm{d}\boldsymbol{x}) = \sum_{i=1}^m \omega_i \boldsymbol{I}(\boldsymbol{x}_i, \boldsymbol{\beta}).$$

We consider models with information matrix $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{\beta}) = Q\big(\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}\big)\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T$, where Q is the intensity function (cf. Fedorov (1972, p. 39)), $\boldsymbol{f}$ is a vector of known regression functions, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T \in \mathbb{R}^p$ is the vector of parameters. This kind of information occurs in a natural way for generalized linear models but also for various other models, like censored data as considered by Konstantinou et al. (2014). Here we consider a model with information matrices of the form

$$\boldsymbol{M}(\xi,\boldsymbol{\beta}) = \sum_{i=1}^{m} \omega_i Q\big(\boldsymbol{f}(\boldsymbol{x}_i)^T\boldsymbol{\beta}\big)\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)^T. \tag{2.1}$$

An optimal design maximizes (or minimizes) some real-valued function of the information matrix with respect to the design. In particular for D-optimality, a design $\xi^* = \xi^*_{\boldsymbol{\beta}}$ with regular information matrix $\boldsymbol{M}(\xi^*_{\boldsymbol{\beta}},\boldsymbol{\beta})$ is locally D-optimal at $\boldsymbol{\beta}$, if $\det\big(\boldsymbol{M}(\xi^*_{\boldsymbol{\beta}},\boldsymbol{\beta})\big) \geq \det\big(\boldsymbol{M}(\xi,\boldsymbol{\beta})\big)$ holds for all $\xi \in \Xi$ (cf. Silvey (1980, p. 54)), where Ξ denotes the set of all probability measures on $\mathscr{X}$. For verifying the D-optimality of a design we make use of an adapted version of the celebrated Kiefer-Wolfowitz equivalence theorem:

Theorem 1. *A design ξ^* is D-optimal if and only if*

$$Q\big(\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}\big) \cdot \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{M}(\xi^*,\boldsymbol{\beta})^{-1}\boldsymbol{f}(\boldsymbol{x}) \leq p \tag{2.2}$$

for all $\boldsymbol{x} \in \mathscr{X}$. At the support points of ξ^* there is equality.

In the following we consider the multilinear case, where $\boldsymbol{f}(\boldsymbol{x}) = (1, \boldsymbol{x}^T)^T$ and $\boldsymbol{x} = (x_1, \ldots, x_{p-1})^T \in \mathscr{X} \subset \mathbb{R}^{p-1}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

Lemma 1. *Let the design region $\mathscr{X}$ be a multidimensional polytope and let the information matrices be of the form* (2.1) *with non-negative function Q. The maximum of the function $d(\boldsymbol{x}) = Q\big(\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}\big)\cdot\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{A}\boldsymbol{f}(\boldsymbol{x})$ with positive definite matrix $\boldsymbol{A}$ is attained only at the edges of the design region.*

Proof. Let $\boldsymbol{A} = (a_{i,j})_{i,j=1,\dots,p}$ and let $\boldsymbol{A}_{11}$ be the submatrix of $\boldsymbol{A}$ formed by deleting the first row and the first column. We have

$$h(\boldsymbol{x}) := \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{A}\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}_{11}\boldsymbol{x} + 2\boldsymbol{a}^T\boldsymbol{x} + a_{1,1}$$

with $\boldsymbol{a} = (a_{1,2},\dots,a_{1,p})^T$. Since $\boldsymbol{A}$ is positive definite, so is $\boldsymbol{A}_{11}$ and hence the function h is strictly convex. For arbitrary $\eta\in\mathbb{R}$ we consider the function d on the hyperplane $H_\eta = \big\{\boldsymbol{x}\in\mathbb{R}^{p-1} : \boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta} = \eta\big\}$. We have $Q\big(\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}\big) = Q(\eta)$ on H_η and $d|_{H_\eta}$ is maximized at the vertices of $\mathscr{X}\cap H_\eta$ because of the strict convexity of the function h. It follows that d is maximized at the edges of $\mathscr{X}$. □

From the proof it follows that for an arbitrary design region $\mathscr{X}$ the function d is maximal at the boundary of the design region. Such functions also occur in the equivalence theorems for other optimality criteria. We directly obtain the following theorem, which gives a general result for a multiple regression model.

Theorem 2. *Let the assumptions of Lemma 1 hold. If the condition in the equivalence theorem is of the form*

$$Q\big(\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}\big)\cdot\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{A}\boldsymbol{f}(\boldsymbol{x}) \le c \tag{2.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

with positive definite matrix $\boldsymbol{A}$ *and constant* $c > 0$*, then the support points of an optimal design are located at the edges of the design region.*

In Theorem 2 the matrix $\boldsymbol{A} = \boldsymbol{A}(\xi^*)$ and the constant $c = c(\xi^*)$ may depend on the optimal design ξ^*. For example, for A-optimality we have $\boldsymbol{A} = \boldsymbol{M}(\xi^*, \boldsymbol{\beta})^{-2}$ and $c = \mathrm{tr}\big(\boldsymbol{M}(\xi^*, \boldsymbol{\beta})^{-1}\big)$. Theorem 2 is valid for many optimality criteria such as the general class of ϕ_p-criteria of Kiefer (1974), including D-optimality.

Examples of optimal designs with this property for multiple regression models can be found in Russell et al. (2009) for the Poisson model and in Kabera et al. (2015) for the logistic regression model with two covariates. Theorem 2 can be extended to unbounded design regions, polyhedra which are defined as the intersection of a finite number of half-spaces. Here a useful example is a quadrant, and then the support points of an optimal design are located on the axes.

Corollary 1. *Let the design region* $\mathscr{X}$ *be a multidimensional polyhedron and let the condition in the equivalence theorem be of the form* (2.3)*. A design* ξ^* *is optimal if and only if* (2.3) *is satisfied on the edges of the polyhedron. If an optimal design exists, then its support points are located at the edges of the design region.*

3. D-optimal designs

In this section we consider a multiple regression model with $p-1$ covariates, $p \geq 3$, and a rectangular design region. The information matrices are assumed to

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

be of the form (2.1). We show that, under some conditions, the D-optimal design in the overall model with $p-1$ covariates can be constructed from the D-optimal designs in the marginal models with a single covariate, where $\tilde{\boldsymbol{f}}_i(x_i) = (1, x_i)^T$. The two-dimensional parameter vector in the i-th marginal model is denoted by $\tilde{\boldsymbol{\beta}}_i$, $i = 1, \ldots, p-1$.

Theorem 3. *Suppose that*

$$\xi_i^* = \begin{Bmatrix} x_i^* & 0 \\ 1/2 & 1/2 \end{Bmatrix}$$

is a D-optimal design in the marginal model with a single covariate, parameter vector $\tilde{\boldsymbol{\beta}}_i = (\beta_0, \beta_i)^T$ and design region $\mathscr{X}_i = (-\infty, 0]$, $i = 1, \ldots, p-1$. Let $\boldsymbol{x}_i^*$ be the embedding of x_i^* in the $(p-1)$-dimensional design region $\mathscr{X} = \mathscr{X}_1 \times \ldots \times \mathscr{X}_{p-1}$ with components $x_{ij}^* = 0$ for $j \neq i$ and $x_{ii}^* = x_i^*$. With $\mathbf{0}_{p-1} = (0, \ldots, 0)^T$ the $(p-1)$-dimensional zero vector, the design

$$\xi^* = \begin{Bmatrix} \boldsymbol{x}_1^* & \ldots & \boldsymbol{x}_{p-1}^* & \mathbf{0}_{p-1} \\ 1/p & \ldots & 1/p & 1/p \end{Bmatrix}$$

is D-optimal in the overall model with $p-1$ covariates and design region $\mathscr{X}$.

The proof is given in the Appendix. The setup for the next result is as follows. Let $S_1, S_2 \subseteq \{1, \ldots, p-1\}$ be index sets (not necessarily non-empty) with $S_1 \cap S_2 = \emptyset$ and $S_1 \cup S_2 = \{1, \ldots, p-1\}$. For $i \in S_1$ and $j \in S_2$ let

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

$\mathscr{X}_i = (-\infty, a_i]$ and $\mathscr{X}_j = [a_j, \infty)$ be the design regions in the marginal models with a single covariate, with $\tilde{\boldsymbol{\beta}}_k = \left(\beta_0 + \sum_{l \neq k} \beta_l a_l, \beta_k\right)^T$ the parameters for the marginal models, $k = 1, \ldots, p-1$, respectively.

Theorem 4. *Suppose that*

$$\xi_i^* = \begin{Bmatrix} x_i^* & a_i \\ 1/2 & 1/2 \end{Bmatrix} \qquad \text{and} \qquad \xi_j^* = \begin{Bmatrix} a_j & x_j^* \\ 1/2 & 1/2 \end{Bmatrix}$$

are D-optimal designs in the marginal models, $i \in S_1$, $j \in S_2$. For $k = 1, \ldots, p-1$, take $\boldsymbol{x}_k^*$ by $x_{kl}^* = a_l$ for $l \neq k$ and $x_{kk}^* = x_k^*$, and let $\boldsymbol{a} = (a_1, \ldots, a_{p-1})^T$. Then the design

$$\xi^* = \begin{Bmatrix} \boldsymbol{x}_1^* & \ldots & \boldsymbol{x}_{p-1}^* & \boldsymbol{a} \\ 1/p & \ldots & 1/p & 1/p \end{Bmatrix} \tag{3.1}$$

is D-optimal in the overall model with $p-1$ covariates and the design region $\mathscr{X} = \mathscr{X}_1 \times \ldots \times \mathscr{X}_{p-1}$.

Proof. With the transformation $z_i = x_i - a_i$ for $i \in S_1$ and $z_j = a_j - x_j$ for $j \in S_2$ the design problem can be reduced to a canonical version, where $z_k \in (-\infty, 0]$, $k = 1, \ldots, p-1$ (see Ford et al. (1992)). The parameter vector $\boldsymbol{\beta}$ is transformed to $\boldsymbol{\beta}_z = \left(\beta_0 + \sum_{l=1}^{p-1} \beta_l a_l, \beta_1, \ldots, \beta_{p-1}\right)^T$. The marginal models can be transformed in the same way. The parameter vectors of the marginal models are transformed to $\tilde{\boldsymbol{\beta}}_{z,k} = \left(\beta_0 + \sum_{l=1}^{p-1} \beta_l a_l, \beta_k\right)^T$ for $k = 1, \ldots, p-1$. The D-optimal designs in

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

the transformed marginal models are given by

$$\xi^*_{z,i} = \begin{Bmatrix} x_i^* - a_i & 0 \\ 1/2 & 1/2 \end{Bmatrix} \qquad \text{and} \qquad \xi^*_{z,j} = \begin{Bmatrix} a_j - x_j^* & 0 \\ 1/2 & 1/2 \end{Bmatrix}$$

for $i \in S_1$ and $j \in S_2$. By Theorem 3 we obtain the D-optimal design for the canonical model. Back transformation yields the D-optimal design (3.1). □

Remark 1. Theorem 4 can also be formulated for bounded design regions. The design regions of the marginal models are given by $\mathscr{X}_i = (-\infty, v_i]$ for $i \in S_1$ and $\mathscr{X}_j = [u_j, \infty)$ for $j \in S_2$, but let $\mathscr{X} = [u_1, v_1] \times \ldots \times [u_{p-1}, v_{p-1}]$ be bounded. Let $a_i = v_i$ for $i \in S_1$ and $a_j = u_j$ for $j \in S_2$. With this definition the vectors $\boldsymbol{a}$ and $\boldsymbol{x}_k^*$, $k = 1, \ldots, p-1$, can be chosen as in Theorem 4. Then the design (3.1) is D-optimal, if $u_i \leq x_i^*$ for $i \in S_1$ and $x_j^* \leq v_j$ for $j \in S_2$.

Example 1. We consider the logistic regression model, for which the function Q is given by $Q(\theta) = e^\theta/(1+e^\theta)^2$. The D-optimal design for the unrestricted design region $\mathscr{X}_i = \mathbb{R}$ in the marginal model with a single covariate and parameter vector $\tilde{\boldsymbol{\beta}}_i = (0, 1)^T$ has two equally weighted support points $x_1^* = -1.543$ and $x_2^* = 1.543$. For the restricted design region $\mathscr{X}_i = [0, \infty)$ the D-optimal design has the two equally weighted support points $x_1^* = 0$ and $x_2^* = 2.399$ (cf. Ford et al. (1992)). In the latter case Theorem 4 is applicable. For the model with two covariates, parameter vector $\boldsymbol{\beta} = (0, 1, 1)^T$ and design region $\mathscr{X} = [0, \infty) \times [0, \infty)$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

we obtain the D-optimal design:

$$\xi^* = \begin{Bmatrix} (2.399, 0) & (0, 2.399) & (0, 0) \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}.$$

This is in agreement with the results of Kabera et al. (2015). The extension to an arbitrary number of covariates is straightforward.

4. Further Applications

We consider models with information matrices of the form (2.1). The intensity function Q is assumed to satisfy the following conditions (cf. Konstantinou et al. (2014)).

(A1) $Q(\theta)$ is positive for all $\theta \in \mathbb{R}$ and twice continuously differentiable.

(A2) $Q'(\theta)$ is positive for all $\theta \in \mathbb{R}$.

(A3) The second derivative $g''(\theta)$ of the function $g(\theta) = 1/Q(\theta)$ is injective.

(A4) The function $Q(\theta)/Q'(\theta)$ is an increasing function.

Condition (A4) is equivalent to $Q(\theta)$ being a log-concave function. Let the function $\phi_{\boldsymbol{a}}(\cdot) : \mathbb{R} \to \mathbb{R}$ with $\boldsymbol{a} = (a_1, \dots, a_{p-1})^T$ be defined as

$$\phi_{\boldsymbol{a}}(x) := x - 2 \cdot \frac{Q\big(\boldsymbol{f}(\boldsymbol{a})^T\boldsymbol{\beta} - x\big)}{Q'\big(\boldsymbol{f}(\boldsymbol{a})^T\boldsymbol{\beta} - x\big)}.$$

Lemma 2. *If (A1), (A2) and (A4) hold, $\phi_{\boldsymbol{a}}$ is strictly increasing, continuous and one-to-one.*

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

For the case of one covariate Konstantinou et al. (2014) determined the D-optimal design for a model with similar conditions on the function Q. With Theorem 4 we can derive the D-optimal design for $p-1$ covariates with arbitrary $p \geq 3$. Here $\boldsymbol{e}_i$ denotes the i-th standard unit vector.

Theorem 5. *Let $\mathscr{X} = [u_1, v_1] \times \ldots \times [u_{p-1}, v_{p-1}]$ and (A1)-(A4) be satisfied for a model with information matrices of the form* (2.1)*. Let $a_i = v_i$ if $\beta_i > 0$ and $a_i = u_i$ if $\beta_i < 0$ for $i = 1, \ldots, p-1$, with $\boldsymbol{a} = (a_1, \ldots, a_{p-1})^T$.*
If $\phi_{\boldsymbol{a}}^{-1}(0) \leq |\beta_i|\,(v_i - u_i)$ for $i = 1, \ldots, p-1$, the design

$$\xi^* = \left\{ \begin{matrix} \boldsymbol{x}_1^* & \boldsymbol{x}_2^* & \ldots & \boldsymbol{x}_p^* \\ 1/p & 1/p & \ldots & 1/p \end{matrix} \right\}$$

with support points $\boldsymbol{x}_i^* = \boldsymbol{a} - (\phi_{\boldsymbol{a}}^{-1}(0)/\beta_i)\,\boldsymbol{e}_i$, $i = 1, \ldots, p-1$, and $\boldsymbol{x}_p^* = \boldsymbol{a}$ is D-optimal.

Proof. For $\beta_i > 0$ the D-optimal design in the marginal model with parameter vector $\tilde{\boldsymbol{\beta}}_i = \big(\beta_0 + \sum_{k \neq i} \beta_k a_k, \beta_i\big)^T$ and design region $\mathscr{X}_i = (-\infty, v_i]$ has two equally weighted support points x_1^* and $x_2^* = v_i$, where x_1^* is the unique solution of (cf. Konstantinou et al. (2014))

$$\beta_i \cdot (v_i - x_1^*) - 2 \cdot \frac{Q\big(\beta_0 + \sum_{k \neq i} \beta_k a_k + \beta_i x_1^*\big)}{Q'\big(\beta_0 + \sum_{k \neq i} \beta_k a_k + \beta_i x_1^*\big)} = 0.$$

With $x_1^* = v_i - z^*$ the equation is given by $\phi_{\boldsymbol{a}}(\beta_i z^*) = 0$ and hence we have $x_1^* = v_i - \phi_{\boldsymbol{a}}^{-1}(0)/\beta_i$.

For $\beta_i < 0$ the D-optimal design in the marginal model with parameter vector $\tilde{\boldsymbol{\beta}}_i = \left(\beta_0 + \sum_{k \neq i} \beta_k a_k, \beta_i\right)^T$ and design region $\mathscr{X}_i = [u_i, \infty)$ has the support points $x_1^* = u_i$ and $x_2^* = u_i - \phi_{\boldsymbol{a}}^{-1}(0)/\beta_i$ with equal weights.

The inequalities $\phi_{\boldsymbol{a}}^{-1}(0) \leq |\beta_i|\,(v_i - u_i)$ for $i = 1, \ldots, p-1$ ensure in both cases that the support points are located inside the design region. Theorem 5 follows from Theorem 4 and Remark 1. $\square$

The statement of Theorem 5 is only valid if all $\beta_1, \ldots, \beta_{p-1}$ are different from zero. If one of the β_i is zero, there might be no optimal design with minimal support. Optimal designs can then be generated incorporating invariance considerations as a product of a uniform design on the vertices for those components where $\beta_k = 0$ and an optimal design as constructed above for the components with non-zero β_j. For the latter components the support points may slightly differ from the optimal marginal design points occurring in Theorem 5, as will be illustrated in Example 3.

Theorem 5 states that one support point of the D-optimal design is the vertex $\boldsymbol{a}$, which depends on the sign of the parameters $\beta_1, \ldots, \beta_{p-1}$. If all these parameters are positive, then $\boldsymbol{a} = \boldsymbol{v} = (v_1, \ldots, v_{p-1})^T$. The other $p-1$ support points are located on the $p-1$ edges that are incident to the vertex $\boldsymbol{a}$ with distance $\phi_{\boldsymbol{a}}^{-1}(0)/\beta_i$ to $\boldsymbol{a}$. The vertex $\boldsymbol{a}$ is the point in the design region with the highest value for the intensity Q. The design region need not be bounded as long

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

as the vertex $\boldsymbol{a}$ remains finite.

In order to calculate the D-optimal design, only $\phi_{\boldsymbol{a}}^{-1}(0)$ has to be determined, so only one equation must be solved. For $Q(\theta) = e^{\theta}$ we obtain $\phi_{\boldsymbol{a}}^{-1}(0) = 2$ and thus the result of Russell et al. (2009) for the Poisson model. For $Q(\theta) = e^{\theta}/(e^{\theta} + \lambda)$ with some constant λ, which corresponds to the intensity function of a negative binomial or Poisson-Gamma model, we get

$$\phi_{\boldsymbol{a}}^{-1}(0) = 2 + W\left(\frac{2}{\lambda} \cdot e^{\boldsymbol{f}(\boldsymbol{a})^T \boldsymbol{\beta} - 2}\right).$$

Here W denotes the principal branch of the Lambert W function, the inverse function of $g(w) = we^{w}$ for $w \geq -1$ (cf. Corless et al. (1996)).

Example 2. For the Poisson-Gamma model Graßhoff et al. (2016) determined D-optimal designs for the case of two binary covariates. They showed that under a certain condition the design with equally weighted binary design points $(1, 0)$, $(0, 1)$ and $(0, 0)$ is D-optimal. For the parameter vector $\boldsymbol{\beta} = (4, -4, -4)^T$, for example, and $\lambda = 1$, this design is also D-optimal on the continuous design region $\mathscr{X} = [0, \infty) \times [0, \infty)$ by Theorem 5. The D-efficiency of the product-type design with equal weights on all four binary design points $(0, 0)$, $(1, 0)$, $(0, 1)$ and $(1, 1)$ is given by 0.772. Hence the D-optimal three-point design performs much better than the product-type design.

Example 3. In order to illustrate the case of some β_i, $i > 0$, being zero, we

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

consider the Poisson model with three covariates and $\boldsymbol{\beta} = (0, -1, -1, 0)^T$. The numerically calculated D-optimal design on the design region $\mathscr{X} = [0, 10]^3$ is given by the product design

$$\begin{aligned} \xi^* &= \xi_{12}^* \otimes \xi_3^* \\ &= \left\{ \begin{matrix} (0,0,0) & (1.86,0,0) & (0,1.86,0) & (0,0,10) & (1.86,0,10) & (0,1.86,10) \\ 0.23 & 0.13 & 0.13 & 0.23 & 0.13 & 0.13 \end{matrix} \right\}, \end{aligned}$$

where ξ_{12}^* assigns weight 0.46 to the vertex $(0, 0)$ and weight 0.27 to axial points $(0, 1.86)$ and $(1.86, 0)$ for the first two components and the uniform design ξ_3^* which assigns equal weights 0.5 to the endpoints 0 and 10 of the design region for the third component. Compared to the product design

$$\xi_{12} \otimes \xi_3^* = \left\{ \begin{matrix} (0,0,0) & (2,0,0) & (0,2,0) & (0,0,10) & (2,0,10) & (0,2,10) \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{matrix} \right\},$$

which is generated from the optimal marginal design ξ_{12} for $\beta_1 = \beta_2 = -1$ on the first two components with equal weights $1/3$ on the vertex $(0, 0)$ and the axial points $(0, 2)$ and $(2, 0)$, and the optimal design ξ_3^* for $\beta_3 = 0$ on the third component, the support points differ slightly whereas there is a larger change in the weights. In fact, for the optimal marginal design ξ_{12}^* the axial points move toward the vertex which itself gets a larger weight. However, the loss in efficiency is less. The product design $\xi_{12} \otimes \xi_3^*$ generated from the optimal marginals still has

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

a relative D-efficiency of 0.965. Further computations suggest that the design ξ_{12}^* is independent of the design region for the third component.

5. Proportional hazards model

In time-to-event experiments, the time duration until the occurrence of some event of interest is observed. The event of interest may be death or cure of individuals under study, or failure of a machine. A typical feature of such experiments is censoring, which occurs when the event of interest is not observed until the end of the experiment.

Let $Y_1, \ldots, Y_n$ be independent, nonnegative random variables, representing the survival times of n individuals and let $C_1, \ldots, C_n$ be the censoring times of the individuals. If the survival time for the i-th individual is greater than its censoring time C_i, then the survival time will be right-censored at C_i. We observe the pairs (T_i, δ_i), where $T_i = \min(Y_i, C_i)$ and δ_i is a censoring indicator with $\delta_i = 1$ if $Y_i \leq C_i$ and $\delta_i = 0$ if $Y_i > C_i$.

We consider type I and random censoring. For type I censoring all individuals join the experiment at the same time and the experiment is terminated at a fixed time point c, so $C_i = c > 0$ for $i = 1, \ldots, n$. For random censoring the censoring times are random variables, which are assumed to be independent of the survival times.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

The Cox proportional hazards model relates the survival times to covariates $\boldsymbol{x}_i$. We assume a constant baseline hazard function $\lambda_0(t) = \lambda = \exp(\beta_0) > 0$ such that

$$\lambda(t; \boldsymbol{x}_i) = \exp\big(\boldsymbol{f}(\boldsymbol{x}_i)^T \boldsymbol{\beta}\big),$$

where $\lambda(t; \boldsymbol{x}_i)$ is the hazard function for the i-th individual under condition $\boldsymbol{x}_i$, which is constant over time, $\boldsymbol{f} = (1, f_1, \ldots, f_{p-1})^T$ is a p-dimensional vector of known regression functions of the covariates and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T \in \mathbb{R}^p$ is the vector of unknown parameters. The survival times Y_i are exponentially distributed (cf. Duchateau and Janssen (2008, pp. 21-22)):

$$Y_i \sim \operatorname{Exp}\Big(e^{\boldsymbol{f}(\boldsymbol{x}_i)^T \boldsymbol{\beta}}\Big), \qquad i = 1, \ldots, n.$$

The Fisher information matrix is given by (cf. Cox and Oakes (1984, p. 82))

$$\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{\beta}) = E_{\boldsymbol{\beta}}(T_i) e^{\boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta}} \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{f}(\boldsymbol{x})^T,$$

which can be written in the form

$$\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{\beta}) = Q\big(\boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta}\big) \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{f}(\boldsymbol{x})^T$$

with intensity function Q. For n independent observations the Fisher information matrix is given by

$$\boldsymbol{I}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{I}(\boldsymbol{x}_i, \boldsymbol{\beta}).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

For type I censoring the function Q is given by $Q(\theta) = 1 - \exp\bigl(-c\exp(\theta)\bigr)$ (Konstantinou et al. (2014)). It satisfies (A1)-(A4).

For random censoring let f_C be the probability density function of the censoring times. Like Konstantinou et al. (2014) we conclude that

$$\begin{aligned} E_{\boldsymbol\beta}(T_i) &= E_{\boldsymbol\beta}\bigl(E_{\boldsymbol\beta}(T_i \mid C_i)\bigr) = \int_0^\infty E_{\boldsymbol\beta}(T_i \mid C_i = c)\cdot f_C(c)\,\mathrm{d}c \\ &= \int_0^\infty \frac{1 - e^{-ce^{\boldsymbol f(\boldsymbol x)^T\boldsymbol\beta}}}{e^{\boldsymbol f(\boldsymbol x)^T\boldsymbol\beta}}\cdot f_C(c)\,\mathrm{d}c \end{aligned}$$

and hence the function Q is given by

$$Q(\theta) = \int_0^\infty \left(1 - e^{-ce^\theta}\right)\cdot f_C(c)\,\mathrm{d}c.$$

It follows that Q is positive. Proofs of the following are given in the Appendix.

Lemma 3. *Let the probability density function f_C be continuous. Then the function Q for random censoring satisfies (A1) and (A2).*

Lemma 4. *Let the probability density function f_C be continuous and log-concave. Then the function Q for random censoring satisfies (A4).*

Many probability distributions have log-concave density functions. Bagnoli and Bergstrom (2005) have compiled a list of probability distributions with log-concave density functions. Truncated distributions may also be of interest for censoring. The truncated distribution of a probability distribution with log-concave density also has a log-concave density (cf. Bagnoli and Bergstrom (2005)). Our results are thus valid for a broad class of models with random censoring.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

Konstantinou et al. (2014) considered censoring times uniformly distributed on the interval $[0, c]$. The resulting intensity function is given by $Q(\theta) = 1 + \left[\exp\bigl(-c\exp(\theta)\bigr) - 1\right] / \bigl(c\exp(\theta)\bigr)$. It can be shown to satisfy (A3), so it satisfies (A1)-(A4). If the censoring times are assumed to be exponentially distributed with parameter $\lambda > 0$, then $Q(\theta) = e^{\theta}/(e^{\theta}+\lambda)$, which corresponds to the intensity function of a negative binomial model. It can easily be shown that this function satisfies (A3) and hence (A1)-(A4). For all these models Theorem 5 gives the D-optimal designs in the case of multiple covariates.

6. Discussion

In this paper we showed that for a large class of models with an arbitrary number of covariates the D-optimal designs can be constructed from the D-optimal designs in the marginal models with a single covariate. The necessary condition is that the D-optimal designs in the marginal models are two-point designs containing a boundary point of the design region as support point. This condition is often satisfied, when the design region is limited such that one of the support points of the D-optimal design on the extended design region $\mathbb{R}$ is located outside of the design region $\mathscr{X}_i$, that is for truncated design regions. For further examples see Biedermann et al. (2006).

The shape of the D-optimal designs is a consequence of a general result, which states that the support points of an optimal design lie on the edges of the

doi:10.5705/ss.202015.0385

design region. It is necessary that only linear terms of the covariates appear in the model. If the model contains interaction terms, this result is no longer true and examples can be found, where a support point must be located inside the design region. Optimal designs with minimal support offer some advantages. In particular for large numbers of variables they can be more easily run in practice than designs with more support points, for example the 2^{p-1} full factorial design based on the optimal marginal design points. It is also easier to round to an exact design if necessary.

Our results may facilitate the search for optimal designs for multiple regression models, and may extend to other optimality criteria. A different approach to the computation of locally optimal designs is weighted designs (cf. Atkinson et al. (2007, Chap.18)), where a prior distribution for the parameters is assumed. These provide a way to overcome the parameter dependence of the locally optimal design. Another possibility is the computation of maximin efficient designs.

Acknowledgements

We would like to thank the referees for their constructive and helpful comments.

Appendix

Proof of Theorem 3. Let $\boldsymbol{x}^* = (x_1^*, \ldots, x_{p-1}^*)^T$, $\mathbf{1}_{p-1} = (1, \ldots, 1)^T$, and $\mathbf{0}_{p-1} = (0, \ldots, 0)^T$ denote $(p-1)$-dimensional vectors. The information matrix of the design ξ^* is given by $\boldsymbol{M}(\xi^*, \boldsymbol{\beta}) = \boldsymbol{X}^T \boldsymbol{Q} \boldsymbol{X}/p$ with

$$\boldsymbol{X} = \begin{pmatrix} 1 & \mathbf{0}_{p-1}^T \\ \mathbf{1}_{p-1} & \operatorname{diag}(\boldsymbol{x}^*) \end{pmatrix}$$

and $\boldsymbol{Q} = \operatorname{diag}\big(Q(\beta_0), Q(\beta_0 + \beta_1 x_1^*), \ldots, Q(\beta_0 + \beta_{p-1} x_{p-1}^*)\big)$, where $\operatorname{diag}(\boldsymbol{x})$ is the diagonal matrix with entries equal to the components of $\boldsymbol{x}$. Let $d(\boldsymbol{x})$ denote the left-hand side of (2.2) for the design ξ^*. By Corollary 1 and Theorem 1 the design ξ^* is D-optimal if and only if for all $x \leq 0$

$$\begin{aligned} p \geq d(x\boldsymbol{e}_i) &= Q\big(\boldsymbol{f}(x\boldsymbol{e}_i)^T \boldsymbol{\beta}\big) \cdot \boldsymbol{f}(x\boldsymbol{e}_i)^T \boldsymbol{M}(\xi^*, \boldsymbol{\beta})^{-1} \boldsymbol{f}(x\boldsymbol{e}_i) \\ &= p \cdot Q(\beta_0 + \beta_i x) \cdot (1, x\boldsymbol{e}_i^T) \boldsymbol{X}^{-1} \boldsymbol{Q}^{-1} (\boldsymbol{X}^T)^{-1} (1, x\boldsymbol{e}_i^T)^T \end{aligned}$$

for $i = 1, \ldots, p-1$. Here $\boldsymbol{e}_i$ denotes the i-th standard unit vector. We note that

$$\boldsymbol{X}^{-1} = \begin{pmatrix} 1 & \mathbf{0}_{p-1}^T \\ -\boldsymbol{y}^* & \operatorname{diag}(\boldsymbol{y}^*) \end{pmatrix}$$

with $\boldsymbol{y}^* = (1/x_1^*, \ldots, 1/x_{p-1}^*)^T$. We have $(1, x\boldsymbol{e}_i^T)\boldsymbol{X}^{-1} = \big(1 - x/x_i^*, (x/x_i^*)\boldsymbol{e}_i^T\big)$ and hence

$$d(x\boldsymbol{e}_i) = p \cdot Q(\beta_0 + \beta_i x) \cdot \left(\frac{\left(1 - \frac{x}{x_i^*}\right)^2}{Q(\beta_0)} + \frac{\left(\frac{x}{x_i^*}\right)^2}{Q(\beta_0 + \beta_i x_i^*)} \right) = p \cdot \frac{1}{2} \cdot d_{M,i}(x),$$

where $d_{M,i}(x)$ denotes the left-hand side of (2.2) for the design ξ_i^* in the marginal model. Since ξ_i^* is D-optimal, we have $d_{M,i}(x) \leq 2$ for all $x \leq 0$ and thus $d(x\boldsymbol{e}_i) \leq p$, which proves the D-optimality of ξ^*. $\square$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

Proof of Lemma 3. The integrand $\big[1-\exp\big(-c\exp(\theta)\big)\big]\cdot f_C(c)$ is twice differentiable with respect to θ and its derivatives are dominated by the integrable function $M\cdot f_C(c)$, where M is a sufficiently large constant. By Lebesgue's Dominated Convergence Theorem, differentiation (with respect to θ) and integration (with respect to c) may be interchanged. Hence Q is twice differentiable and

$$Q'(\theta)=\int_0^\infty ce^\theta e^{-ce^\theta}\cdot f_C(c)\,\mathrm{d}c.$$

Since $c\exp(\theta)\exp\big(-c\exp(\theta)\big)>0$ for $c>0$, we have $Q'(\theta)>0$. □

Proof of Lemma 4. First, we show that the function $m(\theta,c)=1-\exp\big(-c\exp(\theta)\big)$ is log-concave in θ and c. For this purpose we compute the Hessian of $\log m(\theta,c)$. The entries of the symmetric Hessian $\boldsymbol{H}=(h_{ij})_{i,j=1,2}$ are given by:

$$\boldsymbol{H}=\frac{e^\theta e^{-ce^\theta}}{\left(1-e^{-ce^\theta}\right)^2}\begin{pmatrix} c\left(1-ce^\theta-e^{-ce^\theta}\right) & 1-ce^\theta-e^{-ce^\theta}\\ 1-ce^\theta-e^{-ce^\theta} & -e^\theta\end{pmatrix}.$$

The inequality $\exp(x)>1+x$ for $x\neq 0$ yields $1-c\exp(\theta)-\exp\big(-c\exp(\theta)\big)<0$ when $x=-c\exp(\theta)$. Hence $-h_{11}>0$ and

$$\det(-\boldsymbol{H})=\frac{-\left(e^\theta e^{-ce^\theta}\right)^2\cdot\left(1-ce^\theta-e^{-ce^\theta}\right)\cdot\left(1-e^{-ce^\theta}\right)}{\left(1-e^{-ce^\theta}\right)^4}>0.$$

Thus all leading principal minors of $-\boldsymbol{H}$ are positive. It follows that $\boldsymbol{H}$ is negative definite, which proves the log-concavity of $m(\theta,c)$. Since $\log\big(m(\theta,c)\cdot f_C(c)\big)=\log m(\theta,c)+\log f_C(c)$, the product $m(\theta,c)\cdot f_C(c)$ is also log-concave. By Theorem 6 of Prékopa (1973) it follows that $Q(\theta)$ is a log-concave function. □

References

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory* **26**, 445–469.

Biedermann, S., Dette, H., and Zhu, W. (2006). Optimal designs for dose-response models with restricted design spaces. *J. Amer. Statist. Assoc.* **101**, 747–759.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586–602.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996). On the Lambert W Function. *Advances in Computational Mathematics* **5**, 329–359.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.

Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. Springer, New York.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Ford, I., Torsney, B., and Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. R. Statist. Soc. B* **54**, 569–583.

Graßhoff, U., Holling, H., and Schwabe, R. (2016). Optimal Design for the Rasch Poisson-Gamma Model. In *mODa 11 - Advances in Model-Oriented Design and Analysis* (Edited by J. Kunert, C. H. Müller, and A. C. Atkinson), 133–141. Springer.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0385

Kabera, G. M., Haines, L. M., and Ndlovu, P. (2014). The analytic construction of D-optimal designs for the two-variable binary logistic regression model without interaction. *Statistics* **49**, 1169–1186.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849–879.

Konstantinou, M., Biedermann, S., and Kimber, A. (2014). Optimal designs for two-parameter nonlinear models with application to survival models. *Statistica Sinica* **24**, 415–428.

Prékopa, A. (1973). On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum* **34**, 335–343.

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

Russell, K. G., Woods, D. C., Lewis, S. M., and Eccleston, E. C. (2009). D-optimal designs for Poisson regression models. *Statistica Sinica* **19**, 721–730.

Schmidt, D., and Schwabe, R. (2015). On optimal designs for censored data. *Metrika* **78**, 237–257, Erratum, 259.

Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

Institut für Mathematische Stochastik, Otto-von-Guericke-Universität Magdeburg, PF 4120, 39016 Magdeburg, Germany

E-mail: dennis.schmidt@ovgu.de

E-mail: rainer.schwabe@ovgu.de