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

Dennis Schmidt   Rainer Schwabe

November 6, 2015

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.

Keywords multiple regression model · D-optimality · censored data · proportional hazards model · generalized linear models

Mathematics Subject Classification (2000) 62K05 · 62N01

1. Introduction

In order to estimate the parameters of a model as precisely as possible, the optimal choice of control variables is needed, that is, optimal designs have to be computed. In this paper we determine D-optimal designs for a large class of models with an arbitrary number of covariates. We show that they can be constructed from the

D. Schmidt (✉) · R. Schwabe
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

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 many models such as the Poisson model and the negative binomial model. For the Poisson model Russell et al. (2009) determined D-optimal designs, where the number of covariates is arbitrary. This result will follow 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. 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. Since the models considered in this paper are nonlinear, the optimal designs depend on the unknown parameters. Designs, which are optimal for a prespecified parameter value, are called locally optimal (Chernoff, 1953).

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 a certain 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 $0 \leq \omega_i \leq 1$ 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 defined by (cf. Silvey, 1980, p. 53)

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

where $\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{\beta}) = Q\big(\boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta}\big)\, \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{f}(\boldsymbol{x})^T$ is the information matrix at $\boldsymbol{x}$, Q is the intensity function (cf. Fedorov, 1972, p. 39) and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^T \in \mathbb{R}^p$

is the vector of parameters. An optimal design maximizes (or minimizes) some real-valued function of the information matrix with respect to the design.
In particular for the concept of 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 will make use of an adapted version of the celebrated Kiefer-Wolfowitz equivalence theorem:

Theorem 2.1 *A design ξ^* is D-optimal if and only if*

$$Q\big(\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}\big) \cdot \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{M}(\xi^*, \boldsymbol{\beta})^{-1} \boldsymbol{f}(\boldsymbol{x}) \leq p \tag{2.1}$$

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

Throughout this paper we consider a model with information matrices of the form

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \sum_{i=1}^{m} \omega_i Q\big(\boldsymbol{f}(\boldsymbol{x}_i)^T\boldsymbol{\beta}\big) \boldsymbol{f}(\boldsymbol{x}_i) \boldsymbol{f}(\boldsymbol{x}_i)^T \tag{2.2}$$

with $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,p-1})^T \in \mathscr{X} \subset \mathbb{R}^{p-1}$ and $\boldsymbol{f}(\boldsymbol{x}) = (1, \boldsymbol{x}^T)^T$.

Lemma 2.2 *Let the design region $\mathscr{X}$ be a multidimensional polytope and let the information matrices be of the form* (2.2) *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 $\boldsymbol{A} = (a_{i,j})_{i,j=1,\ldots,p}$ 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 $\boldsymbol{a} = (a_{1,2}, \ldots, a_{1,p})^T$. 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. It is valid for many optimality criteria such as the general class of ϕ_p-criteria of Kiefer (1974), which include D-optimality.

Theorem 2.3 *Let the assumptions of Lemma 2.2 hold. If the condition in the equivalence theorem is of the form*

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

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

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 2.3 can be extended to unbounded design regions, that is to polyhedra which are defined as the intersection of a finite number of half-spaces.

Corollary 2.4 *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 condition* (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 rectangular design region. The information matrices are assumed to be of the form (2.2). The following theorem states, that under the given 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 $\tilde{\boldsymbol{\beta}}_i$, $i=1,\dots,p-1$.

Theorem 3.1 *For* $i=1,\dots,p-1$ *let*

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

be 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* $\mathscr{X}_i=(-\infty,0]$*. 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* $x_{ii}^*=x_i^*$*. Let* $\mathbf{0}_{p-1}=(0,\dots,0)^T$ *denote the* $(p-1)$*-dimensional zero vector. Then the design*

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

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

The proof is given in the appendix. The next theorem gives the D-optimal designs for the general case, where the design regions of the marginal models may be left-unbounded or right-unbounded, that is $\mathscr{X}_i = (-\infty, a_i]$ and $\mathscr{X}_j = [a_j, \infty)$ for some $i, j \in \{1, \dots, p-1\}$.

Theorem 3.2 *Let $S_1, S_2 \subseteq \{1, \dots, p-1\}$ be index sets (not necessarily non-empty) with $S_1 \cap S_2 = \emptyset$ and $S_1 \cup S_2 = \{1, \dots, p-1\}$. For $i \in S_1$ and $j \in S_2$ let $\mathscr{X}_i = (-\infty, a_i]$ and $\mathscr{X}_j = [a_j, \infty)$ be the design regions in the marginal models with a single covariate. Moreover, define by $\tilde{\boldsymbol{\beta}}_k = \left(\beta_0 + \sum_{l \neq k} \beta_l a_l, \beta_k\right)^T$ the parameters for the marginal models, $k = 1, \dots, p-1$, respectively. Let*

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

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

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

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 $z_k \in (-\infty, 0]$, $k = 1, \dots, p-1$ (see Ford et al., 1992). The parameter vector $\boldsymbol{\beta}$ is transformed to $\boldsymbol{\beta}_z = \left(\beta_0 + \sum_{l=1}^{p-1} \beta_l a_l, \beta_1, \dots, \beta_{p-1}\right)^T$. 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 $k = 1, \dots, p-1$. The D-optimal designs in 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 3.1 we obtain the D-optimal design for the canonical model. Back transformation yields the D-optimal design (3.1). □

Remark 3.3 Theorem 3.2 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 $\boldsymbol{x}_k^*$, $k = 1, \dots, p-1$, can be chosen as in Theorem 3.2. 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 3.4 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 3.2 is applicable. For the model with two covariates, parameter vector $\boldsymbol{\beta} = (0,1,1)^T$ and design region $\mathscr{X} = [0,\infty) \times [0,\infty)$ we obtain the following 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

As in the last section we consider a model with information matrices of the form (2.2). The intensity function Q is assumed to satisfy the following four 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.

We note that 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

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

The following lemma provides some important properties of the function $\phi_{\boldsymbol{a}}$.

Lemma 4.1 *Let assumptions (A1), (A2) and (A4) be satisfied. The function $\phi_{\boldsymbol{a}}$ is strictly increasing, continuous and one-to-one. Hence the inverse function $\phi_{\boldsymbol{a}}^{-1}$ exists with the same properties.*

For the case of one covariate Konstantinou et al. (2014) have determined the D-optimal design for a model with similar conditions on the function Q. With Theorem 3.2 we can derive the D-optimal design for $p-1$ covariates with arbitrary $p \geq 3$.

Theorem 4.2 *Let $\mathscr{X} = [u_1, v_1] \times \ldots \times [u_{p-1}, v_{p-1}]$ and assumptions (A1)-(A4) be satisfied for a model with information matrices of the form* (2.2)*. Let $a_i = v_i$ if $\beta_i > 0$ and $a_i = u_i$ if $\beta_i < 0$ for $i = 1, \ldots, p-1$. We define $\boldsymbol{a} = (a_1, \ldots, a_{p-1})^T$. If $\phi_{\boldsymbol{a}}^{-1}(0) \leq |\beta_i|\,(v_i - u_i)$ holds for $i = 1, \ldots, p-1$, the design*

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

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

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 the equation (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 $\tilde{\boldsymbol{\beta}}_i = \big(\beta_0 + \sum_{k \neq i} \beta_k a_k, \beta_i\big)^T$ 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 that the support points are located inside the design region. Theorem 4.2 follows from Theorem 3.2 and Remark 3.3. □

Theorem 4.2 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 $\boldsymbol{a}$. We note that 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 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 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, which is defined as the inverse function of $g(w) = we^w$ for $w \geq -1$ (cf. Corless et al., 1996).

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 $Y_1, \dots, Y_n$ be independent, nonnegative random variables, which represent the survival times of the n individuals and let $C_1, \dots, C_n$ 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 the censoring times are fixed and equal for all individuals, that is, $C_i = c > 0$ for $i = 1, \dots, n$. For random censoring the censoring times are random variables, which are assumed to be independent of the survival times.

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

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

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

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

We can write it 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

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

For type I censoring the function Q is given by $Q(\theta) = 1 - \exp\big(-c\exp(\theta)\big)$ (Konstantinou et al., 2014). It satisfies the four conditions (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

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

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. The following two lemmas give further properties of the function Q. The proofs are given in the appendix.

Lemma 5.1 *Let the probability density function f_C be continuous. Then the function Q for random censoring satisfies the conditions (A1) and (A2).*

Lemma 5.2 *Let the probability density function f_C be continuous and log-concave. Then the function Q for random censoring satisfies the condition (A4).*

It turns out that many probability distributions have log-concave density functions. Bagnoli and Bergstrom (2005) have compiled a list of probability distributions with log-concave density functions, which include the normal distribution, logistic distribution, uniform distribution, exponential distribution and Chi-squared distribution and for certain parameter values the Weibull distribution, gamma distribution and beta distribution. 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.
Konstantinou et al. (2014) considered censoring times that are uniformly distributed on the interval $[0, c]$. The resulting intensity function is given by $Q(\theta) = 1 + \big[\exp\big(-c\exp(\theta)\big) - 1\big] / \big(c\exp(\theta)\big)$. It can be shown to satisfy condition (A3), so it satisfies (A1)-(A4). If the censoring times are assumed to be exponentially distributed with parameter $\lambda > 0$, then the function Q is given by $Q(\theta) = e^\theta/(e^\theta + \lambda)$, 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 4.2 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).

Our result was applied to a general class of nonlinear models, which include the proportional hazards model with type I and random censoring as well as the Poisson and negative binomial model. Our assumptions on the intensity function are satisfied for many censoring distributions. They guarantee that the D-optimal designs in the marginal models always are two-point designs, which contain a boundary point of the design region. In order to explicitly calculate the design, only the unique solution of a nonlinear equation must be found.

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 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.

Our results may facilitate the search for optimal designs for multiple regression models. The extension of the present results to other optimality criteria might be a topic for further research. A different approach to the computation of locally optimal designs are 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, which maximize the minimal efficiency with respect to the set of parameters.

A. Appendix

Proof of Theorem 3.1.
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

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

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 (2.1) for the design ξ^*. By Corollary 2.4 and Theorem 2.1 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

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

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

Proof of Lemma 5.1.
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

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

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

Proof of Lemma 5.2.
First, we show that the function $m(\theta, c) = 1 - \exp\bigl(-c\exp(\theta)\bigr)$ 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 $1 - c\exp(\theta) - \exp\bigl(-c\exp(\theta)\bigr) < 0$ 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 $\log\bigl(m(\theta, c) \cdot f_C(c)\bigr) = \log m(\theta, c) + \log f_C(c)$, 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. $\square$

References

Atkinson, A. C., Donev, A. N., 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.

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., Eccleston, E. C. (2009). D-optimal designs for Poisson regression models. *Statistica Sinica*, 19, 721-730.

Schmidt, D., 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.