Optimal Designs for Censored Data

Dennis Schmidt Rainer Schwabe

October 23, 2013

Abstract

In time to event experiments the individuals under study are observed to experience some event of interest. If this event is not observed until the end of the experiment, censoring occurs, which is a common feature in such studies. We consider the proportional hazards model with type I and random censoring and determine locally D- and c-optimal designs for a larger class of nonlinear models with two parameters, where the experimental conditions can be selected from a finite discrete design region as is often the case in practice. Additionally, we compute D-optimal designs for a three-parameter model on a continuous design region.

Keywords censored data · D-optimality · c-optimality · proportional hazards model · discrete design region

Mathematics Subject Classification (2000) 62K05 · 62N01

1. Introduction

The aim of an experiment is to estimate the model parameters as precisely as possible in order to draw the most accurate conclusions from the observations. For this purpose optimal designs are needed since the quality of the parameter estimates depends on the choice of experimental conditions. Optimal designs also ensure cost minimisation in experiments.

A typical feature of time to event experiments is censoring, which can occur in many ways. Konstantinou et al. (2011) considered the proportional hazards model with different censoring schemes and they computed optimal designs for a wider class of models for a continuous design region. However, in practice covariates can often only be chosen

D. Schmidt (✉) · R. Schwabe
Institut für Mathematische Stochastik,
Otto-von-Guericke-Universität Magdeburg,
PF 4120, 39016 Magdeburg, Germany
E-Mail: dennis.schmidt@st.ovgu.de
E-Mail: rainer.schwabe@ovgu.de

from a prespecified discrete set. In drug administration, for example, there is a prespecified set of discrete dose levels for the drugs. We therefore compute optimal designs for a discrete design region extending the existing results for the continuous design region. Section 2 provides the basics of censored data, the model considered in this paper and corresponding information matrices. In section 3 we compute D-optimal designs for a more general model with two parameters containing the proportional hazards model. In the fourth section we will be concerned with determining c-optimal designs for the effect parameter. Section 5 presents an outlook to D-optimal designs for a three parameter model with a continuous design region. The optimal designs depend on the unknown parameters because the model is nonlinear. They are called locally optimal (Chernoff, 1953).

2. Model Specifications

Let $Y_1, Y_2, \ldots, Y_n$ be independent, nonnegative random variables representing the survival times of the individuals. The censoring times of the n individuals will be denoted by $C_1, C_2, \ldots, C_n$. If the event of interest for the ith individual has not occurred until C_i, then its survival time will be right-censored at C_i.
Like Konstantinou et al. (2011) we consider type I and random censoring, when the experiment is terminated at a fixed time point c. For type I censoring all individuals join the experiment at the same time and the censoring times $C_i = c > 0$ are fixed and equal for all individuals. For random censoring the individuals join the experiment at random times $Z_i \in [0, c]$, where the random variables Z_i are assumed to be independent of the survival times Y_i. Hence the censoring times $C_i = c - Z_i$ are random and independent of the survival times Y_i.

One of the most popular models for survival data is the Cox proportional hazards model, which models the survival times of the individuals depending on covariates $\boldsymbol{x}_i$. We investigate the proportional hazards model with a constant baseline hazard function $\lambda_0(t) = \lambda > 0$. By use of the parameterisation $\lambda = \exp(\beta_0)$, the condition $\lambda > 0$ is automatically satisfied and we obtain the model

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

where $\lambda(t; \boldsymbol{x}_i)$ is the constant hazard function for the ith individual under condition $\boldsymbol{x}_i$, $\boldsymbol{f} = (1, f_1, \ldots, f_{p-1})^T$ 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. It follows that the survival times Y_i are exponentially distributed (cf. Duchateau and Janssen, 2008):

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

For type I censoring the Fisher information matrix is given by (cf. Cox and Oakes, 1984; Konstantinou et al., 2011):

$$\boldsymbol{I}(\boldsymbol{x},\boldsymbol{\beta}) = \left(1 - \mathrm{e}^{-c\mathrm{e}^{\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}}}\right) \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T.$$

For random censoring we assume the admission times Z_i to be independent and uniformly distributed on the interval $[0, c]$. Then the same holds for the censoring times $C_i = c - Z_i$. The Fisher information matrix is given by (Konstantinou et al., 2011):

$$\boldsymbol{I}(\boldsymbol{x},\boldsymbol{\beta}) = \left(1 + \frac{\mathrm{e}^{-c\mathrm{e}^{\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}}} - 1}{c\mathrm{e}^{\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}}}\right) \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})^T.$$

For n independent observations, the Fisher information matrix is given by:

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

The Fisher information matrix depends on the control variables the experimenter has to choose. The inverse of the Fisher information matrix is proportional to the asymptotic covariance of the asymptotically efficient maximum likelihood estimator. Our aim is to find the optimal choice of control variables to obtain the most precise estimates of the parameters. This means that we want to maximize the information matrix in a certain sense. The most popular criterion for this is D-optimality, which aims at maximizing the determinant of the information matrix, which can be interpreted to be equivalent to minimizing the volume of the asymptotic confidence ellipsoid for the estimator. We make use of the concept of approximate designs

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

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 reflecting the relative frequency of the corresponding value of the control variable. We have $0 \le \omega_i \le 1$ for $i = 1, \dots, m$ and $\sum_{i=1}^{m} \omega_i = 1$. Such a design can be represented by a probability measure on $\mathscr{X}$. The information matrix $\boldsymbol{M}(\xi,\boldsymbol{\beta})$ of a design ξ is defined by (cf. Silvey, 1980):

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

For the determination of optimal designs real-valued functions of the information matrix are optimised with respect to the design. We consider two of these optimality criteria, D- and c-optimality, which are introduced in the next two sections.

3. D-optimal designs for models with two parameters

Let Ξ be the set of all probability measures on $\mathscr{X}$. A design ξ^* with regular information matrix $\boldsymbol{M}(\xi^*, \boldsymbol{\beta})$ is D-optimal, if $\det\bigl(\boldsymbol{M}(\xi^*, \boldsymbol{\beta})\bigr) \geq \det\bigl(\boldsymbol{M}(\xi, \boldsymbol{\beta})\bigr)$ holds for all $\xi \in \Xi$ (cf. Schwabe, 1996).
Throughout this section we consider a model with information matrices of the form

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \sum_{i=1}^{m} \omega_i Q(\theta_i) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix} \tag{3.1}$$

with positive efficiency function Q, where $\theta_i = \beta_0 + \beta_1 x_i$. For the choice $\boldsymbol{f}(x) = (1, x)^T$ the information matrices for the proportional hazards model with type I censoring and random censoring are both of the form (3.1). We will need the following assumptions on the function Q:

(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.

Both functions $Q(\theta) = 1 - \mathrm{e}^{-ce^{\theta}}$ for type I censoring and $Q(\theta) = 1 + \frac{\mathrm{e}^{-ce^{\theta}} - 1}{ce^{\theta}}$ for random censoring satisfy these four conditions. Konstantinou et al. (2011) considered the same model with assumptions that differ only slightly from ours. Their results remain valid under our assumptions. They showed that the D-optimal design for an interval design region $\mathscr{X} = [u, v]$ is unique and has two equally weighted support points under the assumptions (A1)-(A3). They also determined the D-optimal design, if assumptions (A1)-(A4) are satisfied. For notational reasons we introduce for $\beta_1 \neq 0$ the function

$$\phi(z) = z + \frac{2}{\beta_1} \cdot \frac{Q(\beta_0 + \beta_1 z)}{Q'(\beta_0 + \beta_1 z)} \tag{3.2}$$

which will be helpful for the characterization of the optimal designs in Theorem 3.6 and Theorem 3.8. The following lemma gives some properties of the function ϕ.

Lemma 3.1
Let assumptions (A1), (A2) and (A4) be satisfied. The function ϕ is strictly increasing, continuous and one-to-one. Hence the inverse function ϕ^{-1} exists with the same properties.

For $\beta_1 > 0$ one support point of the D-optimal design for $\mathscr{X} = [u, v]$ is located at the upper boundary $x_2^* = v$ of the design region and the other support point is given by $x_1^* = \max(\phi^{-1}(v), u)$. For $\beta_1 < 0$ the D-optimal design has the support points $x_1^* = u$ and $x_2^* = \min(\phi^{-1}(u), v)$. As these D-optimal designs are saturated, i. e. the number of support points equals the number of parameters, they are uniform on the support with equal weights $w_1^* = w_2^* = 1/p = 1/2$.

In the following, we are interested in finding D-optimal designs for a discrete design region $\mathscr{X} = \{x_1, \ldots, x_k\}$, $x_1 < \ldots < x_k$. Let us first mention a few important results from optimal design theory. A D-optimal design requires two support points in order to have a regular information matrix. By Carathéodory's theorem, for the present situation of $p = 2$ parameters, there exists a D-optimal design with at most three support points (cf. Fedorov, 1972). Hence we can restrict ourselves to consider designs with two and three support points.
For designs with two support points we will make use of the well-known result mentioned above that the D-optimal weights for a design with minimum support are equal (cf. Silvey, 1980). Thus we only have to determine the support points. For most of our results we will make use of an adapted version of the celebrated Kiefer-Wolfowitz equivalence theorem:

Theorem 3.2
A design ξ^* is D-optimal if and only if $\mathrm{tr}\big(\boldsymbol{M}(\xi^*, \boldsymbol{\beta})^{-1}\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{\beta})\big) \leq p$ for all $\boldsymbol{x} \in \mathscr{X}$. At the support points of ξ^* there is equality.

Note that in the present situation $p = 2$. The equivalence theorem is of central importance in optimal design theory providing a necessary and sufficient optimality condition for a design. The following theorem gives the D-optimal design for the case $\beta_1 = 0$.

Theorem 3.3
Let $\mathscr{X} = \{x_1, \ldots, x_k\}$ be a discrete design region with $x_1 < x_2 < \ldots < x_k$. If $\beta_1 = 0$, the unique D-optimal design is given by:

$$\xi^* = \begin{Bmatrix} x_1 & x_k \\ 1/2 & 1/2 \end{Bmatrix}.$$

Proof.
Because $\boldsymbol{M}(\xi, \boldsymbol{\beta}) = Q(\beta_0) \int \boldsymbol{f}\boldsymbol{f}^T \mathrm{d}\xi$ the present information matrix is proportional to the corresponding linear regression setup, and the D-optimal design is supported by the two extremal design points x_1 and x_k. □

In fact this argument also applies to a continuous design region $\mathscr{X} = [u, v]$ and we obtain both boundary points u and v as the optimal support points. From now on let $\beta_1 \neq 0$. As already mentioned, there exists a D-optimal design with at most three support points. If the design region is an interval, then according to Konstantinou et al. (2011) always a two-point design is D-optimal. For a discrete design region this result is no longer true since it is possible that no two-point design is D-optimal but a three-point design as demonstrated by the following example.

Example 3.4
We consider the proportional hazards model with type I censoring and $c = 1$. The design region is chosen as the set $\mathscr{X} = \{0, 1, 2\}$ and let $\beta_0 = -2$ and $\beta_1 = 1.7$ be the true values of the parameters. Table 3.1 shows the values of the determinant of the information matrix for the three possibly optimal two-point designs and a three-point

design ξ^*. The design ξ^* achieves a higher value of the determinant compared to the two-point competitors, so a restriction to two-point designs is not possible in the case of a discrete design region. In fact, the design ξ^* can be shown to be D-optimal.

Design	$\det\big(\boldsymbol{M}(\xi,\boldsymbol{\beta})\big)$
$\xi^* = \begin{Bmatrix} 0 & 1 & 2 \\ 0.197 & 0.319 & 0.484 \end{Bmatrix}$	0.1310
$\xi_1 = \begin{Bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{Bmatrix}$	0.0166
$\xi_2 = \begin{Bmatrix} 0 & 2 \\ 1/2 & 1/2 \end{Bmatrix}$	0.1244
$\xi_3 = \begin{Bmatrix} 1 & 2 \\ 1/2 & 1/2 \end{Bmatrix}$	0.1286

Table 3.1: Comparison between the D-optimal design with three support points and the optimal designs with two support points

Since an optimal two-point design has equal weights, application of the equivalence theorem leads to a first characterisation of optimal two-point designs.

Theorem 3.5
Let $\mathscr{X} = \{x_1, \ldots, x_k\}$ be a discrete design region and $x_i, x_j \in \mathscr{X}$ with $i \neq j$. The design ξ^* with support points x_i and x_j and weights $\omega_i = \omega_j = \frac{1}{2}$ is D-optimal if and only if the following inequality holds for all $x \in \mathscr{X} \setminus \{x_i, x_j\}$:

$$Q(\beta_0 + \beta_1 x) \cdot \frac{Q(\beta_0 + \beta_1 x_i)(x_i - x)^2 + Q(\beta_0 + \beta_1 x_j)(x_j - x)^2}{Q(\beta_0 + \beta_1 x_i) Q(\beta_0 + \beta_1 x_j)(x_j - x_i)^2} \leq 1.$$

Proof.
We use the notation $\theta_x = \beta_0 + \beta_1 x$ for all $x \in \mathscr{X}$ and apply the equivalence theorem. The design ξ^* is D-optimal if and only if

$$\operatorname{tr}\big(\boldsymbol{M}(\xi^*, \boldsymbol{\beta})^{-1} \boldsymbol{I}(x, \boldsymbol{\beta})\big) = \frac{2Q(\theta_x) \cdot \big[Q(\theta_{x_i})(x_i - x)^2 + Q(\theta_{x_j})(x_j - x)^2\big]}{Q(\theta_{x_i}) Q(\theta_{x_j})(x_j - x_i)^2} \overset{!}{\leq} 2 \qquad (3.3)$$

for all $x \in \mathscr{X}$. For the support points $x = x_i$ and $x = x_j$ there is equality in (3.3) and division by two completes the proof. □

We did not use the properties of the function Q so far. Applying (A1), (A2) and (A4), the following theorem shows that only two of the $\binom{k}{2}$ designs with two support points are possible as D-optimal designs.

Theorem 3.6
Let $\mathscr{X} = \{x_1, \ldots, x_k\}$ be a discrete design region with $x_1 < x_2 < \ldots < x_k$ and let assumptions (A1), (A2) and (A4) be satisfied.
a) For $\beta_1 > 0$ denote by $x^- = \max\{x_i \in \mathscr{X} : x_i < \tilde{x}\}$ and $x^+ = \min\{x_i \in \mathscr{X} : x_i \geq \tilde{x}\}$, where $\tilde{x} = \phi^{-1}(x_k)$. The design

$$\xi^* = \begin{Bmatrix} x_1^* & x_k \\ 1/2 & 1/2 \end{Bmatrix}$$

maximises the determinant of the information matrix in the class of all two-point designs, where $x_1^* = x^-$ if $Q(\beta_0 + \beta_1 x^-)(x_k - x^-)^2 \geq Q(\beta_0 + \beta_1 x^+)(x_k - x^+)^2$ and $x_1^* = x^+$ for "$\leq$". If equality occurs, there are two optimal designs.
In the case $\{x_i \in \mathscr{X} : x_i < \tilde{x}\} = \emptyset$ we have $x_1^* = x^+ = x_1$.

b) For $\beta_1 < 0$ denote by $x^- = \max\{x_i \in \mathscr{X} : x_i \leq \tilde{x}\}$ and $x^+ = \min\{x_i \in \mathscr{X} : x_i > \tilde{x}\}$, where $\tilde{x} = \phi^{-1}(x_1)$. The design

$$\xi^* = \begin{Bmatrix} x_1 & x_2^* \\ 1/2 & 1/2 \end{Bmatrix}$$

maximises the determinant of the information matrix in the class of all two-point designs, where $x_2^* = x^-$ if $Q(\beta_0 + \beta_1 x^-)(x^- - x_1)^2 \geq Q(\beta_0 + \beta_1 x^+)(x^+ - x_1)^2$ and $x_2^* = x^+$ for "$\leq$". If equality occurs, there are two optimal designs.
In the case $\{x_i \in \mathscr{X} : x_i > \tilde{x}\} = \emptyset$ we have $x_2^* = x^- = x_k$.

The proof is given in the appendix. We now proceed with determining D-optimal designs with three support points. We thus have to compute the optimal weights and support points. Starting with the weights, let us first consider a discrete design region with only three elements since the general case can easily be reduced to this one.

Theorem 3.7
Let $\mathscr{X} = \{x_1, x_2, x_3\}$ be the design region, $d_{ij} = Q(\beta_0 + \beta_1 x_i)Q(\beta_0 + \beta_1 x_j)(x_j - x_i)^2$ for $i, j \in \{1, 2, 3\}$ and $l \in \{1, 2, 3\} \setminus \{i, j\}$.

$$\xi = \begin{Bmatrix} x_i & x_j \\ 1/2 & 1/2 \end{Bmatrix} \text{ is } D\text{-optimal} \iff \frac{d_{il} + d_{jl}}{d_{ij}} \leq 1.$$

Otherwise, if none of the three two-point designs is D-optimal, the unique D-optimal design is given by

$$\xi_4 = \begin{Bmatrix} x_1 & x_2 & x_3 \\ \omega_1 & \omega_2 & \omega_3 \end{Bmatrix}$$

with the following weights

$$\omega_1 = \frac{d_{23}(d_{13} + d_{12} - d_{23})}{d}, \quad \omega_2 = \frac{d_{13}(d_{23} + d_{12} - d_{13})}{d}, \quad \omega_3 = \frac{d_{12}(d_{23} + d_{13} - d_{12})}{d}, \tag{3.4}$$

where $d = d_{12}(d_{23} + d_{13} - d_{12}) + d_{23}(d_{13} + d_{12} - d_{23}) + d_{13}(d_{23} + d_{12} - d_{13})$.

This result concerning the optimal weights can be generalised to an arbitrary discrete design region $\mathscr{X} = \{x_1, \ldots, x_k\}$. Since there exists a D-optimal design ξ^* with three support points, if no two-point design is D-optimal, there are three support points $x_i, x_j, x_l \in \mathscr{X}$ forming the support of ξ^*. Application of Theorem 3.7 to the design region $\{x_i, x_j, x_l\}$ yields the D-optimal design with the optimal weights (3.4) depending on the support points x_i, x_j and x_l. This design is also D-optimal on $\mathscr{X}$.
Having determined the optimal weights, we now turn to the support points.

Theorem 3.8
Let $\mathscr{X} = \{x_1, \ldots, x_k\}$ be a discrete design region with $x_1 < x_2 < \ldots < x_k$ and let assumptions (A1)-(A4) be satisfied. If no design with two support points is D-optimal, then a design with three support points $x_1^* < x_2^* < x_3^*$ is D-optimal. Let x^- and x^+ be defined as in Theorem 3.6 for $\beta_1 > 0$ and $\beta_1 < 0$ respectively.
a) For $\beta_1 > 0$ the D-optimal design has the support points $x_1^* = x^-$, $x_2^* = x^+$ and $x_3^* = x_k$.
b) For $\beta_1 < 0$ the D-optimal design has the support points $x_1^* = x_1$, $x_2^* = x^-$ and $x_3^* = x^+$.

A measure for the goodness of a design is its efficiency. The D-efficiency of a design ξ is defined by (cf. Atkinson and Donev, 1996)

$$\mathrm{eff}_D(\xi, \boldsymbol{\beta}) = \left(\frac{\det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big)}{\det\big(\boldsymbol{M}(\xi^*_{\boldsymbol{\beta}}, \boldsymbol{\beta})\big)} \right)^{\frac{1}{2}},$$

where $\xi^*_{\boldsymbol{\beta}}$ is the locally D-optimal design. The next theorem by Atwood (1969) gives a bound for the loss in case of restriction to designs with two support points.

Theorem 3.9
Let $\mathscr{X} = \{x_1, \ldots, x_k\}$ be a discrete design region. If a three-point design $\xi^*_{\boldsymbol{\beta}}$ is locally D-optimal at $\boldsymbol{\beta}$, then there exists a two-point design ξ' with $\mathrm{eff}_D(\xi', \boldsymbol{\beta}) \geq \frac{\sqrt{3}}{2} \approx 0.866$.

We note that if the function Q is a strictly increasing function, then the inequality in Theorem 3.9 holds strictly.

Remark 3.10
Let assumptions (A1), (A2) and (A4) be satisfied and let $\beta_1 \neq 0$. For all $\varepsilon > 0$ there exists a discrete design region $\mathscr{X}$ such that $\mathrm{eff}_D(\xi', \boldsymbol{\beta}) \leq \frac{\sqrt{3}}{2} + \varepsilon$ holds for all two-point designs ξ' on $\mathscr{X}$.

4. c-optimal designs for two parameters

In this section we determine c-optimal designs for a model with information matrix of the form (3.1). For c-optimal designs interest lies primarily in estimating a particular linear aspect $\boldsymbol{c}^T\boldsymbol{\beta}$ with minimum variance. A design ξ^* is c-optimal, if the linear aspect $\boldsymbol{c}^T\boldsymbol{\beta}$ is identifiable, that is $\boldsymbol{c} = \boldsymbol{M}(\xi,\boldsymbol{\beta})\,\boldsymbol{a}$ for some vector $\boldsymbol{a}$, and if $\boldsymbol{c}^T\boldsymbol{M}(\xi^*,\boldsymbol{\beta})^-\boldsymbol{c} \leq \boldsymbol{c}^T\boldsymbol{M}(\xi,\boldsymbol{\beta})^-\boldsymbol{c}$ holds for all $\xi \in \Xi$ for which $\boldsymbol{c}^T\boldsymbol{\beta}$ is identifiable. The matrix $\boldsymbol{M}(\xi,\boldsymbol{\beta})^-$ is a generalized inverse of $\boldsymbol{M}(\xi,\boldsymbol{\beta})$.
For the binary design region $\mathscr{X} = \{0,1\}$ with a treatment group and a control group the parameter β_1 represents the effect of the treatment on the hazard rate. In dose-response studies the parameter β_1 represents the effect of increasing the dose. Hence we are interested in estimating the effect parameter β_1 as accurately as possible and we treat β_0 as a nuisance parameter. We thus consider $\boldsymbol{c} = (0\;\;1)^T$-optimality.
Konstantinou et al. (2011) showed that a c-optimal design requires two support points and they determined the c-optimal design for an interval design region $\mathscr{X} = [u,v]$. Since for c-optimality there always exists an optimal design with only two support points (cf. Pukelsheim, 1993), there is also a c-optimal two-point design for a discrete design region, contrary to D-optimality. The optimal weights for such a design are given by (Pukelsheim and Torsney, 1991; Konstantinou et al., 2011):

$$\omega_1 = \frac{\sqrt{Q(\beta_0+\beta_1 x_2)}}{\sqrt{Q(\beta_0+\beta_1 x_1)}+\sqrt{Q(\beta_0+\beta_1 x_2)}},\;\; \omega_2 = \frac{\sqrt{Q(\beta_0+\beta_1 x_1)}}{\sqrt{Q(\beta_0+\beta_1 x_1)}+\sqrt{Q(\beta_0+\beta_1 x_2)}}.$$

Theorem 4.1
Let $\mathscr{X} = \{x_1,\ldots,x_k\}$ be the design region with $x_1 < x_2 < \ldots < x_k$. If $\beta_1 = 0$, the design

$$\xi^* = \begin{Bmatrix} x_1 & x_k \\ 1/2 & 1/2 \end{Bmatrix}$$

is c-optimal for β_1.

Proof.
The result follows again directly from the linear case because of the proportionality of the information matrices. □

Analogously to Theorem 3.6 for D-optimality the following theorem holds for c-optimality.

Theorem 4.2
Let $\mathscr{X} = \{x_1,\ldots,x_k\}$ be a discrete design region with $x_1 < x_2 < \ldots < x_k$ and let assumptions (A1), (A2) and (A4) be satisfied. Let the function h be defined as follows:

$$h(z_1,z_2) = \left(\frac{1}{\sqrt{Q(\beta_0+\beta_1 z_1)}}+\frac{1}{\sqrt{Q(\beta_0+\beta_1 z_2)}}\right)^2 \cdot \frac{1}{(z_2-z_1)^2}.$$

a) For $\beta_1 > 0$ let $\tilde{x}$ be the unique solution of the following equation:

$$\beta_1 \cdot (x_k - \tilde{x}) - 2 \cdot \frac{Q(\beta_0 + \beta_1 \tilde{x})}{Q'(\beta_0 + \beta_1 \tilde{x})} \cdot \left(1 + \frac{\sqrt{Q(\beta_0 + \beta_1 \tilde{x})}}{\sqrt{Q(\beta_0 + \beta_1 x_k)}}\right) = 0.$$

Denote by $x^- = \max\{x_i \in \mathscr{X} : x_i < \tilde{x}\}$ and $x^+ = \min\{x_i \in \mathscr{X} : x_i \geq \tilde{x}\}$. The design

$$\xi^* = \left\{ \begin{array}{cc} x_1^* & x_k \\ \frac{\sqrt{Q(\beta_0+\beta_1 x_k)}}{\sqrt{Q(\beta_0+\beta_1 x_1^*)}+\sqrt{Q(\beta_0+\beta_1 x_k)}} & \frac{\sqrt{Q(\beta_0+\beta_1 x_1^*)}}{\sqrt{Q(\beta_0+\beta_1 x_1^*)}+\sqrt{Q(\beta_0+\beta_1 x_k)}} \end{array} \right\}$$

is c-optimal for β_1, where $x_1^* = x^-$ if $h(x^-, x_k) \leq h(x^+, x_k)$ and $x_1^* = x^+$ for "$\geq$". In the case $\{x_i \in \mathscr{X} : x_i < \tilde{x}\} = \emptyset$ we have $x_1^* = x^+ = x_1$.
b) For $\beta_1 < 0$ let $\tilde{x}$ be the unique solution of the following equation:

$$\beta_1 \cdot (\tilde{x} - x_1) + 2 \cdot \frac{Q(\beta_0 + \beta_1 \tilde{x})}{Q'(\beta_0 + \beta_1 \tilde{x})} \cdot \left(1 + \frac{\sqrt{Q(\beta_0 + \beta_1 \tilde{x})}}{\sqrt{Q(\beta_0 + \beta_1 x_1)}}\right) = 0.$$

Denote by $x^- = \max\{x_i \in \mathscr{X} : x_i \leq \tilde{x}\}$ and $x^+ = \min\{x_i \in \mathscr{X} : x_i > \tilde{x}\}$. The design

$$\xi^* = \left\{ \begin{array}{cc} x_1 & x_2^* \\ \frac{\sqrt{Q(\beta_0+\beta_1 x_2^*)}}{\sqrt{Q(\beta_0+\beta_1 x_1)}+\sqrt{Q(\beta_0+\beta_1 x_2^*)}} & \frac{\sqrt{Q(\beta_0+\beta_1 x_1)}}{\sqrt{Q(\beta_0+\beta_1 x_1)}+\sqrt{Q(\beta_0+\beta_1 x_2^*)}} \end{array} \right\}$$

is c-optimal for β_1, where $x_2^* = x^-$ if $h(x_1, x^-) \leq h(x_1, x^+)$ and $x_2^* = x^+$ for "$\geq$". In the case $\{x_i \in \mathscr{X} : x_i > \tilde{x}\} = \emptyset$ we have $x_2^* = x^- = x_k$.

The proof is omitted since it follows analogously to the proof of Theorem 3.6.

5. D-optimal designs for three parameters

In this section we consider an interval design region $\mathscr{X} = [u, v]$. We incorporate an additional quadratic term into the model such that $\boldsymbol{f}(x) = (1, x, x^2)$ and the information matrix is of the form

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \sum_{i=1}^{m} \omega_i Q(\theta_i) \begin{pmatrix} 1 & x_i & x_i^2 \\ x_i & x_i^2 & x_i^3 \\ x_i^2 & x_i^3 & x_i^4 \end{pmatrix},$$

where $\theta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$. A D-optimal design must have at least three support points. The following theorem gives conditions under which the determinant of a three-point design is maximised regardless of the third support point at both boundary points of the design region.

Theorem 5.1
Let $\mathscr{X} = [u,v]$ and assumptions (A1), (A2) and (A4) be satisfied. A D-optimal design in the class of all three-point designs has support points $x_1^* < x_2^* < x_3^*$ with weights $\omega_1 = \omega_2 = \omega_3 = \frac{1}{3}$. Let the function k be defined as follows:

$$\psi(z_1, z_2) = \frac{1}{2}(\beta_1 + 2\beta_2 z_1)(z_2 - z_1) + 2 \cdot \frac{Q(\beta_0 + \beta_1 z_1 + \beta_2 z_1^2)}{Q'(\beta_0 + \beta_1 z_1 + \beta_2 z_1^2)}.$$

a) Let $\beta_2 > 0$ and let the inequality $(-\beta_1/\beta_2) \geq u + v$ hold. Then we have $x_1^* = u$. Additionally, we have $x_3^* = v$, if the equation

$$\psi(x, u) = 0 \tag{5.1}$$

has at most one solution x or has no solution $x \in (u, v)$. If $x_1^* = u$ and $x_3^* = v$, the support point x_2^* is given by the smallest solution $x \in \left(u, \frac{u+v}{2}\right]$ of the following equation:

$$(\beta_1 + 2\beta_2 x)(v - x)(x - u) + 2 \cdot \frac{Q(\beta_0 + \beta_1 x + \beta_2 x^2)}{Q'(\beta_0 + \beta_1 x + \beta_2 x^2)} \cdot (v + u - 2x) = 0. \tag{5.2}$$

b) Let $\beta_2 > 0$ and let the inequality $(-\beta_1/\beta_2) \leq u + v$ hold. Then we have $x_3^* = v$. Additionally, we have $x_1^* = u$, if the equation

$$\psi(x, v) = 0 \tag{5.3}$$

has at most one solution x or has no solution $x \in (u, v)$. If $x_1^* = u$ and $x_3^* = v$, the support point x_2^* is given by the largest solution $x \in \left[\frac{u+v}{2}, v\right)$ of equation (5.2).

c) Let $\beta_2 < 0$ and let the inequality $-\beta_1/(2\beta_2) \leq u$ hold. Then we have $x_1^* = u$. Additionally, we have $x_3^* = v$, if the unique solution $x > u$ of equation (5.1) satisfies $x \geq v$. If $x_1^* = u$ and $x_3^* = v$, the support point x_2^* is given by the unique solution $x \in \left(u, \frac{u+v}{2}\right)$ of equation (5.2).

d) Let $\beta_2 < 0$ and let the inequality $-\beta_1/(2\beta_2) \geq v$ hold. Then we have $x_3^* = v$. Additionally, we have $x_1^* = u$, if the unique solution $x < v$ of equation (5.3) satisfies $x \leq u$. If $x_1^* = u$ and $x_3^* = v$, the support point x_2^* is given by the unique solution $x \in \left(\frac{u+v}{2}, v\right)$ of equation (5.2).

Example 5.2
a) We consider the proportional hazards model with type I censoring and $c = 1$. Let $\boldsymbol{\beta} = (0, 1, 1)^T$ be the parameter vector and let the design region be given by the interval $[0, 1]$. Since $\beta_2 = 1 > 0$ and $(-\beta_1/\beta_2) = -1 \leq 1 = u + v$, part b) of Theorem 5.1 holds and hence $x_3^* = 1$. The function $Q(\theta)/Q'(\theta)$ is an increasing function by (A4) and is thus minimal for $\theta \to -\infty$. Since numerator and denominator converge to zero as $\theta \to -\infty$, application of l'Hôpital's rule yields:

$$\lim_{\theta \to -\infty} \frac{Q(\theta)}{Q'(\theta)} = \lim_{\theta \to -\infty} \frac{Q'(\theta)}{Q''(\theta)} = \lim_{\theta \to -\infty} \frac{c\mathrm{e}^{\theta}\mathrm{e}^{-c\mathrm{e}^{\theta}}}{c\mathrm{e}^{\theta}\mathrm{e}^{-c\mathrm{e}^{\theta}}(1 - c\mathrm{e}^{\theta})} = \lim_{\theta \to -\infty} \frac{1}{1 - c\mathrm{e}^{\theta}} = 1.$$

It follows that $-2Q(\beta_0 + \beta_1 x + \beta_2 x^2)/Q'(\beta_0 + \beta_1 x + \beta_2 x^2) < -2$. For a solution of equation (5.3) to exist it is necessary that the maximum value of the quadratic function $l(x) := \frac{1}{2}(\beta_1 + 2\beta_2 x)(1 - x)$ is greater than 2. The maximum is located at $\tilde{x} = 1/4$ with $l(\tilde{x}) = 9/16$. Hence equation (5.3) has no solution and we have $x_1^* = 0$ by Theorem 5.1. The third support point x_2^* can be calculated by equation (5.2) as $x_2^* = 0.534$.
b) We consider the same model as in a) but with parameter vector $\boldsymbol{\beta} = (0, 4, 1)^T$. As before we have $x_3^* = 1$. Now equation (5.3) has two solutions $x = -0.939$ and $x = -0.416$, which are both located outside the interval $(0, 1)$. Again we get $x_1^* = 0$ and we obtain $x_2^* = 0.500$ by equation (5.2).

6. Discussion

The model underlying this paper is the proportional hazards model with two different types of censoring. We considered type I and random censoring and we investigated a more general model with two parameters including both censoring schemes. For an arbitrary discrete design region the D- and c-optimal designs were found. Our results are also valid for the poisson regression model and extend the results by Rodríguez-Torreblanca and Rodríguez-Díaz (2007).
For D-optimality either a two-point or a three-point design is D-optimal. We determined the optimal design in the class of all two-point designs and identified the optimal weights and support points of the optimal three-point design if no two-point design is D-optimal. For c-optimality there exists an optimal design with two support points. We computed the c-optimal design for the effect parameter.
Since the optimal designs depend on the unknown parameters, an initial guess of the parameters is needed but this could result in a poor design when the parameters are misspecified. If a range for the parameters can be given, standardised maximin optimal designs can be computed, which maximise the minimal efficiency with respect to the set of parameters. For the continuous design region Konstantinou et al. (2011) determined such designs. Extension of these results to a discrete design region might be a topic for further research.

Furthermore, we considered a model with a quadratic term and three parameters, where the design region is an interval. It was shown that under certain conditions the D-optimal design in the class of all three-point designs must have both boundary points of the design region and for this case an optimal design was determined. Since the computations become more extensive for a three parameter model, it becomes more difficult to achieve analytic results and hence numerical methods will gain in importance.

A. Appendix

Proof of Theorem 3.6.
a) The first part of the proof is similar to that of Konstantinou et al. (2011) for the interval design region. An optimal design ξ^* with two support points $x_1^* < x_2^*$ has equal weights. The determinant of the information matrix is given by:

$$\det\big(\boldsymbol{M}(\xi^*, \boldsymbol{\beta})\big) = \frac{1}{4} Q(\beta_0 + \beta_1 x_1^*) Q(\beta_0 + \beta_1 x_2^*)(x_2^* - x_1^*)^2.$$

From assumption (A2) it follows that the function Q is strictly increasing. Since $\beta_1 > 0$, the determinant is increasing with x_2^* and it is maximised for $x_2^* = x_k$. To obtain the second support point we have to maximise the function $r(x) := Q(\beta_0 + \beta_1 x)(x_k - x)^2$ over $x \in \mathscr{X}$. We consider $r(x)$ as a function of real x and show that it has only one extremum in $(-\infty, x_k)$, which is a maximum. The first derivative is

$$r'(x) = \beta_1 Q'(\beta_0 + \beta_1 x)(x_k - x)^2 - 2Q(\beta_0 + \beta_1 x)(x_k - x)$$

with a zero located at $x = x_k$. Since the function Q is positive, so is the function $r(x)$ for all $x \neq x_k$ and hence there is a minimum at $x = x_k$. We may get further zeros of r' by solving the equation $\phi(x) = x_k$. The solution $\tilde{x} = \phi^{-1}(x_k)$ is unique by Lemma 3.1. Since ϕ is strictly increasing and $\phi(x_k) > x_k$, the second zero $\tilde{x}$ is located in the interval $(-\infty, x_k)$. It is a maximum of r, because the derivative changes sign at this point.
It follows that the function $r(x)$ is strictly decreasing to both sides around the maximum and for the discrete design region it is maximised either at x^- or x^+ or eventually at both points. Figure A.1 illustrates the situation.

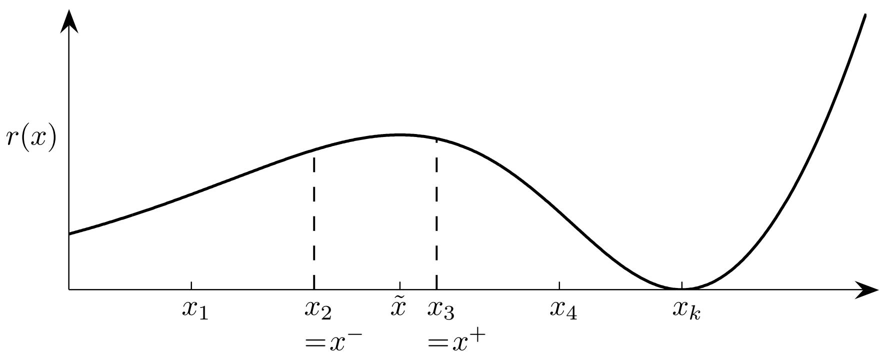

Figure A.1: Sketch of a typical graph for the function $r(x) = Q(\beta_0 + \beta_1 x)(x_k - x)^2$

Since $x_k > \tilde{x}$, we have $\{x_i \in \mathscr{X} : x_i \geq \tilde{x}\} \neq \emptyset$. If $\tilde{x} \leq x_1$, the maximum is located at $\tilde{x} \leq x_1 = x^+$ and in this case $x^+ = x_1$ is the second support point. If $x_1 < \tilde{x}$ then x^- and x^+ are located at two adjacent design points x_i and x_{i+1}, say, x_2^* has to be chosen at that of these two points, at which r attains the maximal value.
b) The case $\beta_1 < 0$ can be treated in a similar way or the proof may be obtained by symmetry considerations. □

Proof of Theorem 3.7.
The optimality conditions for the two-point designs result from Theorem 3.5. If none of these designs is D-optimal, then a D-optimal design must have three support points. The determinant of the information matrix for such a design is given by:

$$\det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big) = \omega_1\omega_2 d_{12} + \omega_1\omega_3 d_{13} + \omega_2\omega_3 d_{23}.$$

Substituting $\omega_3 = 1 - \omega_1 - \omega_2$ and setting the partial derivatives with respect to ω_1 and ω_2 equal to zero leads to the following system of linear equations:

$$\begin{pmatrix} -2d_{13} & d_{12} - d_{13} - d_{23} \\ d_{12} - d_{13} - d_{23} & -2d_{23} \end{pmatrix} \begin{pmatrix} \omega_1^* \\ \omega_2^* \end{pmatrix} = \begin{pmatrix} -d_{13} \\ -d_{23} \end{pmatrix}.$$

Solving this system yields the solution given in (3.4).
We denote the matrix on the left-hand side by $\boldsymbol{A}$ and show that its determinant, which is the denominator of the optimal weights, is not equal to zero. The determinant of $\boldsymbol{A}$ is given by:

$$\det(\boldsymbol{A}) = d_{12}(d_{23} + d_{13} - d_{12}) + d_{23}(d_{13} + d_{12} - d_{23}) + d_{13}(d_{23} + d_{12} - d_{13}).$$

Since no design with two support points is D-optimal, the optimality conditions for the two-point designs in Theorem 3.7 are not satisfied. Hence the three expressions in parentheses are positive and so is the determinant.
Because the design ξ_4 contains the optimal weights, it is the unique D-optimal design. □

Proof of Theorem 3.8.
a) We first prove that a D-optimal design must contain the support point x_k. Let ξ be a design with support points $x_i, x_j, x_l \in \mathscr{X}$, $x_i < x_j < x_l$, and weights ω_i, ω_j and ω_l. The determinant of the information matrix is given by:

$$\det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big) = \omega_i\omega_j d_{ij} + \omega_i\omega_l d_{il} + \omega_j\omega_l d_{jl}. \tag{A.1}$$

As in the proof of Theorem 3.6, we conclude that the determinant is increasing with x_l and it is maximised for $x_l = x_k$.

We will now show that a D-optimal design with three support points $x_i < x_j < x_k$ must have a support point less than $\phi^{-1}(x_k)$. Assume that all three support points were greater than $\phi^{-1}(x_k)$. We apply Theorem 2 of Konstantinou et al. (2011) to the design region $[x_i, x_k]$ and obtain the unique D-optimal design ξ^* with support points x_i and x_k and equal weights. Since it is D-optimal on $[x_i, x_k]$, it is also D-optimal on $\{x_i, x_j, x_k\}$, which is a contradiction.

We continue to show that it is impossible that both support points x_i and x_j are less than $\phi^{-1}(x_k)$. Assume that $x_i < x_j < \phi^{-1}(x_k)$. Since $\phi(x_j) > x_j$, we have $\phi(x_j) \in (x_j, x_k)$. Application of Theorem 2 of Konstantinou et al. (2011) to the design region $[x_i, \phi(x_j)]$ yields the unique D-optimal design ξ^* with support points x_j and $\phi(x_j)$ and equal weights. Since it is D-optimal on $[x_i, \phi(x_j)]$, it is also D-optimal on $\{x_i, x_j, \phi(x_j)\}$ and hence Theorem 3.7 gives the inequality

$$\frac{Q(\theta_i)Q(\theta_j)(x_j - x_i)^2}{Q(\theta_j)Q(\theta_{\phi_j})(\phi(x_j) - x_j)^2} + \frac{Q(\theta_i)Q(\theta_{\phi_j})(\phi(x_j) - x_i)^2}{Q(\theta_j)Q(\theta_{\phi_j})(\phi(x_j) - x_j)^2} \leq 1, \tag{A.2}$$

where $\theta_{\phi_j} = \beta_0 + \beta_1\phi(x_j)$. Let the functions $h_1(x)$ and $h_2(x)$ be defined as follows:

$$h_1(x) = \frac{Q(\theta_i)(x_j - x_i)^2}{Q(\beta_0 + \beta_1 x)(x - x_j)^2}, \qquad h_2(x) = \frac{Q(\theta_i)(x - x_i)^2}{Q(\theta_j)(x - x_j)^2}.$$

The function h_1 is a strictly decreasing function in x. The derivative of $h_2(x)$ is given by:

$$h_2'(x) = \frac{2Q(\theta_i)(x - x_i)(x_i - x_j)}{Q(\theta_j)(x - x_j)^3}.$$

Since $x_i - x_j < 0$, the derivative is negative for $x > x_j$ and hence the function $h_2(x)$ is strictly decreasing for $x > x_j$. Thus the following inequality holds as well, which follows from (A.2) by substituting $\phi(x_j)$ by x_k:

$$\frac{Q(\theta_i)Q(\theta_j)(x_j - x_i)^2}{Q(\theta_j)Q(\theta_k)(x_k - x_j)^2} + \frac{Q(\theta_i)Q(\theta_k)(x_k - x_i)^2}{Q(\theta_j)Q(\theta_k)(x_k - x_j)^2} \leq 1.$$

Theorem 3.7 states that the design with support points x_j and x_k and equal weights is D-optimal on the design region $\mathscr{X} = \{x_i, x_j, x_k\}$, which contradicts our assumption that no two-point design is D-optimal. Hence we have $x_1^* < \phi^{-1}(x_k) < x_2^*$.

It remains to show that $x_1^* = x^-$ and $x_2^* = x^+$. Let $\boldsymbol{M}(\xi, \boldsymbol{\beta})^{-1} = (m_{ij})_{i,j=1,2}$ denote the inverse of the information matrix. By the equivalence theorem, the design ξ is D-optimal if and only if

$$\frac{m_{11} + 2m_{12}x + m_{22}x^2}{2} \leq \frac{1}{Q(\theta)} =: g(\theta)$$

for all $x \in \mathscr{X}$. As $\beta_1 \neq 0$ we can write the left-hand side as a quadratic polynomial in $\theta = \beta_0 + \beta_1 x$ with suitable coefficients $c_1, c_2, c_3 \in \mathbb{R}$. Then the design ξ is D-optimal if and only if $c_1\theta^2 + c_2\theta + c_3 \leq g(\theta)$ for all $\theta \in \beta_0 + \beta_1\mathscr{X}$, which is equivalent to:

$$k(\theta) := g(\theta) - c_1\theta^2 - c_2\theta - c_3 \geq 0 \quad \forall \theta \in \beta_0 + \beta_1\mathscr{X}.$$

We consider again k as a function on $\mathbb{R}$ and we have $k''(\theta) = g''(\theta) - 2c_1$. Since $g''(\theta)$ is injective by (A3), the function $k''(\theta)$ can have at most one zero. From Rolle's theorem it follows that $k'(\theta)$ has at most two zeros. Hence $k(\theta)$ has at most one minimum. If $k(\beta_0 + \beta_1\phi^{-1}(x_k)) \geq 0$, then the three-point design would also be D-optimal on $\mathscr{X} \cup \{\phi^{-1}(x_k)\}$, which is a contradiction because the two-point design with support points $\phi^{-1}(x_k)$ and x_k is the unique D-optimal design.
We thus get $k(\beta_0 + \beta_1\phi^{-1}(x_k)) < 0 = k(\beta_0 + \beta_1 x_1^*) = k(\beta_0 + \beta_1 x_2^*)$. Since $k(\theta)$ has at

most one minimum and $x_1^* < \phi^{-1}(x_k) < x_2^*$, it follows that the D-optimal three-point design must have the support points $x_1^* = x^-$ and $x_2^* = x^+$.
b) For $\beta_1 < 0$ the proof works analogously to the case $\beta_1 > 0$. □

Proof of Remark 3.10.
We give the proof only for the case $\beta_1 > 0$, since for $\beta_1 < 0$ it works in the same way. In the proof of Theorem 3.6 a) we have shown that for arbitrary $a \in \mathbb{R}$ the function $r(x) := Q(\beta_0 + \beta_1 x)(a - x)^2$ has exactly one extremum in the interval $(-\infty, a)$, which is a maximum. Hence $r(x)$ is eventually decreasing for $x \to -\infty$ and bounded below by zero, thus converging to some limit $b \geq 0$. Since the factor $(a - x)^2$ tends to ∞ as $x \to -\infty$, we have $\lim_{x\to-\infty} Q(\beta_0 + \beta_1 x) = 0$.
We now show that $\lim_{x\to-\infty} r(x) = 0$ for fixed $a \in \mathbb{R}$. Assume that $\lim_{x\to-\infty} r(x) = b > 0$. Application of l'Hôpital's rule yields

$$b = \lim_{x\to-\infty} \frac{Q(\beta_0 + \beta_1 x)}{\left(\frac{1}{(a-x)^2}\right)} = \lim_{x\to-\infty} \frac{\beta_1 Q'(\beta_0 + \beta_1 x)}{\left(\frac{2}{(a-x)^3}\right)} = \lim_{x\to-\infty} \frac{1}{2}\beta_1 Q'(\beta_0 + \beta_1 x)(a - x)^3$$

and it follows:

$$1 = \frac{b}{b} = \lim_{x\to-\infty} \frac{Q(\beta_0 + \beta_1 x)(a - x)^2}{\frac{1}{2}\beta_1 Q'(\beta_0 + \beta_1 x)(a - x)^3} = \lim_{x\to-\infty} \frac{Q(\beta_0 + \beta_1 x)}{Q'(\beta_0 + \beta_1 x)} \cdot \frac{2}{\beta_1(a - x)}.$$

The second factor $2/\big(\beta_1(a - x)\big)$ converges to 0 as $x \to -\infty$. Thus the first factor $Q(\beta_0 + \beta_1 x)/Q'(\beta_0 + \beta_1 x)$ has to tend to ∞ as $x \to -\infty$, contradicting assumption (A4) that this function is increasing.

Now a suitable design region $\mathscr{X} = \{x_1, x_2, x_2\}$ is constructed as follows. Let $\varepsilon > 0$ and x_3 may be chosen arbitrarily. The elements $x_1 < \phi^{-1}(x_3)$ and $x_2 \in (\phi^{-1}(x_3), x_3)$ will be chosen to satisfy the following two conditions:

i) $d_{13} = d_{23}$

ii) $\dfrac{d_{13} - d_{12}}{2d_{13} + d_{12}} \leq \dfrac{4}{3}\varepsilon^2$

Since $\lim_{x_1\to-\infty} Q(\beta_0 + \beta_1 x_1)(x_3 - x_1)^2 = 0$, the equation is satisfied for some x_1. The inequality can also be satisfied, because the left-hand side converges to zero as $x_2 \to x_3$. By Theorem 3.6 a) the design ξ' with support points x_1 and x_3 and equal weights 1/2 is an optimal two-point design. Let ξ be the three-point design with support points x_1, x_2 and x_3 and corresponding weights 1/3. The determinant of the information matrix of ξ is given by:

$$\det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big) = \frac{1}{9}d_{12} + \frac{2}{9}d_{13}.$$

The denominator of the left-hand side of inequality ii) is equal to $9 \cdot \det\big(\boldsymbol{M}(\xi, \boldsymbol{\beta})\big)$ and we get the inequality

$$\begin{aligned}\det\bigl(\boldsymbol{M}(\xi,\boldsymbol{\beta})\bigr) &\geq \frac{1}{9}\left[d_{13} - 12\varepsilon^2 \det\bigl(\boldsymbol{M}(\xi,\boldsymbol{\beta})\bigr)\right] + \frac{2}{9}d_{13} \\ &= -\frac{4}{3}\varepsilon^2 \det\bigl(\boldsymbol{M}(\xi,\boldsymbol{\beta})\bigr) + \frac{1}{3}d_{13} \\ &= -\frac{4}{3}\varepsilon^2 \det\bigl(\boldsymbol{M}(\xi,\boldsymbol{\beta})\bigr) + \frac{4}{3}\det\bigl(\boldsymbol{M}(\xi',\boldsymbol{\beta})\bigr),\end{aligned}$$

which is equivalent to $\left(\frac{3}{4}+\varepsilon^2\right)\cdot\det\bigl(\boldsymbol{M}(\xi,\boldsymbol{\beta})\bigr) \geq \det\bigl(\boldsymbol{M}(\xi',\boldsymbol{\beta})\bigr)$. Now we have

$$\mathrm{eff}_D(\xi',\boldsymbol{\beta}) \leq \left(\frac{\det\bigl(\boldsymbol{M}(\xi',\boldsymbol{\beta})\bigr)}{\det\bigl(\boldsymbol{M}(\xi,\boldsymbol{\beta})\bigr)}\right)^{\frac{1}{2}} \leq \left(\frac{3}{4}+\varepsilon^2\right)^{\frac{1}{2}} \leq \frac{\sqrt{3}}{2} + \varepsilon,$$

which proves the assertion. □

Proof of Theorem 5.1.
a) A D-optimal three-point design ξ^* with support points $x_1 < x_2 < x_3$ has equal weights 1/3. The determinant of the information matrix is given by

$$\det\bigl(\boldsymbol{M}(\xi^*,\boldsymbol{\beta})\bigr) = \frac{1}{27}Q(\theta_1)Q(\theta_2)Q(\theta_3)(x_3-x_1)^2(x_3-x_2)^2(x_2-x_1)^2,$$

where $\theta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$ for $i = 1,2,3$ is a quadratic polynomial in x_i and thus symmetric about the minimum at $x_i = -\beta_1/(2\beta_2)$. Hence $Q(\theta_i)$ is also symmetric. Since the function Q is strictly increasing, $Q(\theta_i)$ has a minimum at $x_i = -\beta_1/(2\beta_2)$ and is strictly increasing to both sides. It follows that $Q(\theta_i)$ can only be maximal at one of the boundary points of the interval $[u,v]$. If $-\beta_1/(2\beta_2) \geq (u+v)/2$ holds, then $Q(\beta_0+\beta_1 u+\beta_2 u^2) \geq Q(\beta_0+\beta_1 v+\beta_2 v^2)$ and the determinant of the information matrix is maximised for $x_1 = u$. If $-\beta_1/(2\beta_2) \leq (u+v)/2$ holds, then the determinant is maximised for $x_3 = v$. We consider the first case since the second case follows analogously.
Now we want to maximise the determinant for fixed x_2 over x_3. Therefore we have to maximise the function $h(x_3) := Q(\beta_0+\beta_1 x_3+\beta_2 x_3^2)(x_3-u)^2(x_3-x_2)^2$, which has only two zeros at $x_3 = u$ and $x_3 = x_2$ and is positive elsewhere. Hence $h(x_3)$ is initially increasing for $x_3 > x_2$. If no maximum exists, then $h(x_3)$ is increasing for all $x_3 > x_2$ and is maximised for $x_3 = v$. This is the case if and only if the derivative has at most one zero $\tilde{x}_3 > x_2$. Then $\tilde{x}_3$ is a saddle point of $h(x_3)$ since the function tends to ∞ as $x_3 \to \infty$. If there exist several zeros of the derivative which are located outside the design region, then $h(x_3)$ is also maximised for $x_3 = v$. For $x_3 > x_2$ the derivative of $h(x_3)$ is equal to zero if and only if:

$$l(x_3,x_2) := \frac{(\beta_1+2\beta_2 x_3)(x_3-u)(x_3-x_2)}{2x_3-x_2-u} = -2\cdot\frac{Q(\beta_0+\beta_1 x_3+\beta_2 x_3^2)}{Q'(\beta_0+\beta_1 x_3+\beta_2 x_3^2)} =: r(x_3).$$

Since the function $r(x_3)$ is negative, zeros of $h'(x_3)$ can only exist if $l(x_3,x_2)$ is also negative. The function $l(x_3,x_2)$ is negative only if $x_2 < x_3 < -\beta_1/(2\beta_2)$. The location of the zeros depends on x_2. The derivative of $l(x_3, x_2)$ with respect to x_2 is given by

$$\frac{\mathrm{d}l(x_3, x_2)}{\mathrm{d}x_2} = -\frac{(\beta_1 + 2\beta_2 x_3)(x_3 - u)^2}{(2x_3 - x_2 - u)^2}$$

and is thus positive for $x_3 \in \big(x_2, -\beta_1/(2\beta_2)\big)$. Hence $l(x_3, x_2)$ is minimal on this interval for $x_2 = u$. If the equation $l(x_3, u) = r(x_3)$ has at most one solution, then the equation $l(x_3, x_2) = r(x_3)$ has at most one solution for all $x_2 \in [u, v]$ and $h'(x_3)$ has no second zero for $x_3 > x_2$. This fact is visualised by Figure A.2.

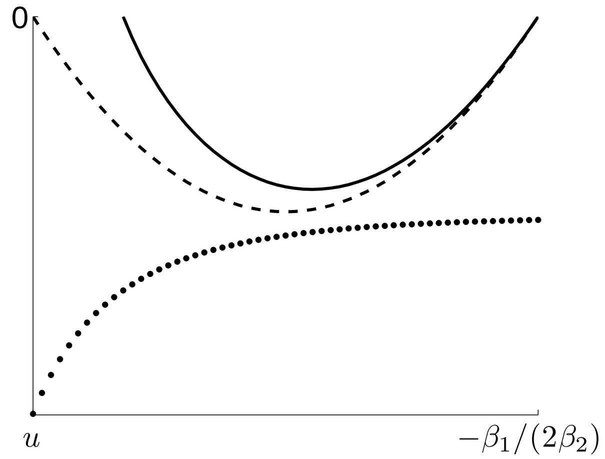

Figure A.2: Proof sketch (dashed line: $l(x_3, u)$, solid line: $l(x_3, x_2)$ for $x_2 > u$, dotted line: $r(x_3)$)

The function $l(x_3, u)$ is given by $l(x_3, u) = \frac{1}{2}(\beta_1 + 2\beta_2 x_3)(x_3 - u)$ and the equation $l(x_3, u) = r(x_3)$ is equivalent to (5.1).

We now consider the case $h'(x_3)$ having several zeros. For fixed $x_2 \in \big(u, -\beta_1/(2\beta_2)\big)$ let a zero of $h'(x_3)$ be located at $\tilde{x}_3$. Since $l(x_3, x_2)$ is minimised for $x_2 = u$, it follows that $l(\tilde{x}_3, u) < l(\tilde{x}_3, x_2) = r(\tilde{x}_3)$. Moreover, we have $\lim_{x_3 \to u} l(x_3, u) = 0 > r(u)$. By continuity of both functions there must exist a point of intersection at $x_3 < \tilde{x}_3$. This shows that the first point of intersection of $l(x_3, x_2)$ and $r(x_3)$ is minimal for $x_2 = u$. If this zero of $h'(x_3)$ is located outside the interval (u, v) for $x_2 = u$, then it is located outside this interval for all $x_2 > u$ and $h(x_3)$ is maximised for $x_3 = v$.

A D-optimal three-point design ξ^* has thus the support points $x_1 = u$ and $x_3 = v$ under the conditions given. For maximisation of the determinant of the information matrix over x_2 we have to maximise the function $k(x_2) := Q(\beta_0 + \beta_1 x_2 + \beta_2 x_2^2)(v - x_2)^2(x_2 - u)^2$. The term $(v - x_2)^2(x_2 - u)^2$ is symmetric about $x_2 = (u + v)/2$. Since the symmetry point of $Q(\beta_0 + \beta_1 x_2 + \beta_2 x_2^2)$ is located at $x_2 = -\beta_1/(2\beta_2) \geq (u + v)/2$, the function $k(x_2)$ is maximised in the interval $\big(u, (u + v)/2\big]$. For $x_2 \in \big(u, (u + v)/2\big]$ the derivative of $k(x_2)$ is equal to zero if and only if:

$$(\beta_1 + 2\beta_2 x_2)(v - x_2)(x_2 - u) + 2 \cdot \frac{Q(\beta_0 + \beta_1 x_2 + \beta_2 x_2^2)}{Q'(\beta_0 + \beta_1 x_2 + \beta_2 x_2^2)} \cdot (v + u - 2x_2) = 0. \quad \text{(A.3)}$$

We distinguish two cases.
Case 1: Let $-\beta_1/(2\beta_2) = (u+v)/2$. Then $k'(x_2)$ has a zero at $x_2 = (u+v)/2$. Further zeros exist if and only if:

$$-\beta_2(v - x_2)(x_2 - u) = -2 \cdot \frac{Q(\beta_0 + \beta_1 x_2 + \beta_2 x_2^2)}{Q'(\beta_0 + \beta_1 x_2 + \beta_2 x_2^2)}.$$

The left-hand side is strictly decreasing for $x_2 \in \big(u, (u+v)/2\big]$. The function Q/Q' is increasing by (A4) and it follows that the right-hand side is increasing on the interval $\big(u, (u+v)/2\big]$. Hence there can be at most one point of intersection. This point cannot be a saddle point since the derivative changes sign. Thus it is a maximum and there is a minimum at $x_2 = (u+v)/2$ because of the symmetry of $k(x_2)$ about $(u+v)/2$. If no point of intersection exists, then the maximum is located at $x_2 = (u+v)/2$. In both cases the maximum is located at the smallest solution of equation (A.3).

Case 2: Let $-\beta_1/(2\beta_2) > (u+v)/2$. Then $k'(x_2)$ has no zero at $x_2 = (u+v)/2$. We have $k'(x_2) = 0$, if and only if:

$$l(x_2) := \frac{(\beta_1 + 2\beta_2 x_2)(v - x_2)(x_2 - u)}{v + u - 2x_2} = -2 \cdot \frac{Q(\beta_0 + \beta_1 x_2 + \beta_2 x_2^2)}{Q'(\beta_0 + \beta_1 x_2 + \beta_2 x_2^2)} =: r(x_2).$$

The derivative of $l(x_2)$ is given by

$$l'(x_2) = \beta_1 + 2\beta_2 x_2 + \frac{2(v - x_2)(x_2 - u)\big(\beta_1 + \beta_2(v+u)\big)}{(v + u - 2x_2)^2}$$

and it is negative for $x_2 \in \big(u, (u+v)/2\big)$. Hence $l(x_2)$ is strictly decreasing on this interval. The function $r(x_2)$ is increasing on this interval. Since $l(u) = 0 > r(u)$ and $l(x_2)$ tends to $-\infty$ as $x_2 \to (u+v)/2$, there exists exactly one point of intersection at which $k(x_2)$ is maximised.
b) For $\beta_2 < 0$ the proof follows with similar arguments. □

References

Atkinson, A. C. and Donev, A. N. (1996). *Optimum Experimental Designs.* Clarendon Press, Oxford.

Atwood, C. L. (1969). *Optimal and efficient designs of experiments.* Ann. Math. Statist., 40, 1570-1602.

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

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.

Konstantinou, M., Biedermann, S. and Kimber, A. (2011). *Optimal designs for two-parameter nonlinear models with application to survival models*. Statistica Sinica (to appear, doi:10.5705/ss.2011.271).

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Pukelsheim, F. and Torsney B. (1991). *Optimal Weights for Experimental Designs on Linearly Independent Support Points*. Ann. Statist., Vol. 19, No. 3, 1614-1625.

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.

Schwabe, R. (1996). *Optimum Designs for Multi-Factor Models*. Springer, New York.

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