UNIVERSITY OF SOUTHAMPTON

Faculty of Human and Social Sciences
Mathematical Sciences

Locally optimal and robust designs for two-parameter nonlinear models with application to survival models

by

Maria Konstantinou

Thesis submitted for the degree of Doctor of Philosophy
September 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

Mathematical Sciences

Doctor of Philosophy

Locally optimal and robust designs for two-parameter nonlinear models with application to survival models

by Maria Konstantinou

Survival experiments are conducted in many industrial and biomedical applications to evaluate the effect of a method or treatment on the time until the occurrence of an event. Thus the survival models used are often two-parameter models and involve data that are subject to censoring, that is, the event of interest is not observed for all the subjects in the experiment. Finding efficient designs for survival experiments is vital in order to minimise their running costs and maximise the precision of their conclusions. The current research incorporates censoring in the well established methodology of Design of Experiments to produce novel methods for planning such experiments.

We provide analytical characterisations of locally D- and c-optimal designs for a wide class of two-parameter nonlinear models that includes many commonly used survival models, based on easily verifiable assumptions. These reduce the numerical effort for design search substantially and can be interpreted directly by practitioners.

In order to overcome the parameter dependence of locally optimal designs we investigate the construction of standardised maximin D- and c-optimal designs and of cluster designs and illustrate our results using the exponential-based proportional hazards model. Different censoring mechanisms are incorporated and the robustness of our designs against parameter misspecifications is verified.

A general framework is set up for the construction of optimal designs for partial likelihood estimation for Cox's proportional hazards model. We show that under Type-I censoring, the designs derived assuming the exponential distribution are optimal for any baseline hazard. We also demonstrate that c-optimal designs for the exponential regression model based on full likelihood, are efficient for partial likelihood estimation.

We also provide analytical characterisations of minimax D- and c-optimal designs that are robust to deviations from the exponential-based proportional hazards model. The latter results coincide with the locally c-optimal designs which can therefore be used even if the exponential distribution assumption is incorrect

Throughout this project we show that traditional designs currently in use are not the best choice in many practical scenarios and we provide efficient alternatives that can be directly implemented by practitioners. These alternatives have the potential to influence the design of future survival experiments.

Contents

1 Introduction **1**

2 Basics **5**

2.1 Optimal experimental planning for parameter estimation 5

2.1.1 Exact and approximate designs 5

2.1.2 Optimality criteria 6

2.1.3 General equivalence and Caratheodory's theorems 7

2.1.4 Bayesian designs 8

2.2 Modelling time to event data 8

2.2.1 Censoring mechanisms 9

2.2.2 Survivor and hazard functions 9

2.2.3 Proportional hazards models 10

2.2.4 Accelerated failure time models 11

2.3 Optimal designs for survival models 12

2.3.1 Type-I censoring 13

2.3.2 Random censoring 13

3 Literature review **15**

3.1 Locally optimal designs 15

3.2 Parameter-robust designs 17

3.2.1 Maximin designs 17

3.2.2 Bayesian designs 18

3.2.3 Cluster designs 19

3.3 Model-robust designs 19

3.4 Optimal designs for survival models 20

3.4.1 Accelerated failure time models 20

3.4.2 Exponential regression model 21

3.4.3 Cox's proportional hazards model 22

4 Locally optimal designs **23**

4.1 Class of models 24

4.1.1 Information matrix assumptions 24
4.1.2 Examples 25
4.2 Locally D-optimal designs 26
4.2.1 The criterion 27
4.2.2 Binary design space 27
4.2.3 Continuous design space 27
4.3 Locally c-optimal designs for estimating β 29
4.3.1 The criterion 29
4.3.2 General results 30
4.3.3 Binary design space 31
4.3.4 Continuous design space 31
4.4 Application to the exponential regression model 33
4.4.1 The model 33
4.4.2 No censoring 34
4.4.3 Right-censoring 34
4.4.4 Locally optimal designs 36
4.4.5 Recommendations 38

5 Parameter-robust designs 39
5.1 Standardised maximin optimal designs 40
5.1.1 Standardised maximin D-optimal designs 41
5.1.2 Standardised maximin c-optimal designs 43
5.2 Cluster designs 45
5.3 Robustness analysis 46
5.3.1 Locally D-optimal designs 47
5.3.2 Locally c-optimal designs 47
5.3.3 Standardised maximin optimal designs 48
5.3.4 Cluster designs 49
5.3.5 Comparison of designs 49

6 Optimal designs for partial likelihood information 53
6.1 Cox's model and partial likelihood function 54
6.2 Optimality criterion 56
6.3 Minimum variance designs using the partial likelihood method 58
6.3.1 No censoring 59
6.3.2 Type-I censoring 60
6.3.3 Random censoring 63
6.4 Comparison of designs arising from full and partial likelihood methods 65
6.4.1 Numerical results 66
6.4.2 Analytical results 68

6.5 Conclusions 69

7 Model-robust designs **71**

7.1 Models and contamination functions 72

7.2 Optimality criteria 73

7.2.1 Preliminaries 73

7.2.2 Mean squared error matrix 76

7.2.3 Minimax designs 78

7.3 Minimax optimal two-point designs for binary design space 78

7.3.1 Minimax c-optimal designs for β 79

7.3.2 Minimax c-optimal designs for α 80

7.3.3 Minimax D-optimal designs 80

7.4 Minimax optimal two-point designs for continuous design space 82

7.4.1 Minimax c-optimal designs for β 82

7.4.2 Minimax c-optimal designs for α 85

7.4.3 Minimax D-optimal designs 86

7.5 Application to Freireich data 87

7.6 Conclusions 89

8 Discussion **91**

8.1 Results and conclusions 91

8.2 Future work 95

A **97**

A.1 Proofs for Chapter 1 97

A.1.1 Statement 1 97

A.1.2 Statement 2 97

A.1.3 Statement 3 99

A.1.4 Statement 4 101

A.1.5 Statement 5 103

A.1.6 Lemma 1 105

A.1.7 Theorem 2(b) 106

A.1.8 Theorem 4(b) 107

B **111**

B.1 Proofs for Chapter 5 111

B.1.1 Statement 1 (Proof of Theorem 5) 111

B.1.2 Statement 2 (Proof of Theorem 5) 112

B.2 Proofs for Chapter 6 113

B.2.1 Proposition 1 113

B.3 Proofs for Chapter 7 115
B.3.1 Theorem 8 115
B.3.2 Theorem 9(b) 116
B.3.3 Theorem 10 117

References **121**

List of Tables

4.1 Support points for some selected locally D-optimal designs for model (4.7) under Type-I censoring 36
4.2 Support points for some selected locally D-optimal designs for model (4.7) under random censoring 36
4.3 Weights on $x = 0$ for some selected locally c-optimal designs for β for model (4.7) under Type-I censoring 37
4.4 Weights on $x = 0$ for some selected locally c-optimal designs for β for model (4.7) under random censoring 37

5.1 D-efficiencies for some selected locally D-optimal designs 47
5.2 Weights for some selected locally c-optimal designs 48
5.3 c-efficiencies for the locally c-optimal designs of Table 5.2 48

6.1 Optimal designs for binary and continuous design spaces and efficiencies, in percent, of the balanced design in the absence of censoring 60
6.2 Optimal weights $1-\omega$ corresponding to $x_2 = 1$ and efficiency, in percent, of the balanced design for a binary design space and Type-I censoring 61
6.3 Support points $\{x_1, x_2\}$, optimal weights $1-\omega$ at point x_2 and efficiency, in percent, of the balanced design under Type-I censoring for $\mathcal{X} = [0, 1]$ 62
6.4 Optimal weights $1-\omega$ corresponding to $x_2 = 1$ and efficiency, in percent, of the balanced design for a binary design space and random censoring 64
6.5 Support points $\{x_1, x_2\}$, optimal weights $1-\omega$ at point x_2 and efficiency, in percent, of the balanced design under random censoring for $\mathcal{X} = [0, 1]$ 65
6.6 Efficiencies, in percent, of full likelihood designs under Type-I censoring for a binary (and a continuous) design space 67
6.7 Efficiencies, in percent, of full likelihood designs under random censoring for a binary (and a continuous) design space 67

List of Figures

5.1 Boxplots of D-efficiencies calculated for 9 designs for 1000 parameter vectors 50

5.2 Boxplots of c-efficiencies calculated for 6 designs for 1000 parameter vectors 50

7.1 Minimax D-optimal weight ω at point 0 for $g \in \mathcal{G}_1$ 88

7.2 Minimax D-optimal weight ω at point 0 for $g \in \mathcal{G}_2$ 88

Declaration of Authorship

I, MARIA KONSTANTINOU, declare that the thesis entitled

LOCALLY OPTIMAL AND ROBUST DESIGNS FOR TWO-PARAMETER NONLINEAR MODELS WITH APPLICATION TO SURVIVAL MODELS

and the work presented in the thesis are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work has been published as: Konstantinou, Biedermann and Kimber (2013).

Signed ....

Date ....

Acknowledgements

First and foremost I would like to thank my supervisors Dr Stefanie Biedermann and Dr Alan Kimber for their patience, guidance and continuous support during the four years of my PhD studies. The knowledge I have gained from them is invaluable.

I would also like to thank the people who for fifteen years I have the honour to call friends for all the wonderful times we have shared. I owe special thanks to Eleni and Christos for always being there, for believing in me and keep reminding me about what are the important things in life.

A big thank you also to Evangelos for his encouragement and understanding during the last and perhaps the most difficult part of my PhD life. Thank you for holding my hand through all the too many stressful days and for the happy and carefree memories you have offered me.

I dedicate this work to two of the most intelligent and strong women I have ever met, my grandmothers Eleni and Maria who have been an inspiration throughout my life.

Finally, I would like to thank my parents Gavrili and Kakia for their faith in me and their unconditional love and support. You were, are and will always be my role models. Thank you for everything.

to my grandmother Eleni and the
loving memory of my grandmother Maria ...

Chapter 1

Introduction

The motivation for this project comes from survival experiments which are of great importance mainly due to their wide range of applications. The models involved in such experiments called survival models arise in almost all areas of scientific research, for example, in medicine, biostatistics, engineering and social sciences. Survival experiments are also vital since they are conducted in order to evaluate the efficacy, the benefits and the safety of existing and new methods used in each of their applications. Designing these experiments in a more efficient way is both interesting and beneficial as this will reduce the amount of resources required for their execution and also improve inference and hence the accuracy of the conclusions drawn at the end of the experiments. Therefore, our work is on the interface of two major areas of statistical research, namely Design of Experiments (DoE) and Survival Analysis, and it combines the well established methodology of the former area with the most important features of the latter.

Survival models which are considered in Survival Analysis are usually nonlinear and their response variable is the time until the occurrence of a particular event, such as the death of a patient or the failure of a machine. The event of interest may also be a non fatal outcome but could be, for example, the cure of a patient. The resulting data are thus often referred to as time to event data or survival data. Their main characteristic is censoring which occurs when the event is not observed for some of the subjects under investigation. This phenomenon results in the standard methods of modelling and analysis to be unsuitable for survival data.

The area of DoE provides us with powerful analytical methods for determining the experimental conditions where measurements must be taken in order for the goal of the experiment to be accomplished. The designs arising from DoE theory are optimal in the sense that they require fewer numbers of subjects to be utilised in the experiment to achieve the same accuracy of conclusions as with a suboptimal design. Hence the use of an optimal design reduces both the experimentation time and cost. Moreover, optimal designs maximise the information contained in the data, which is extremely important since no statistical analysis or modelling technique can extract information the data do not have.

To further stress the importance of optimal experimental design for survival experiments let us consider the example of a clinical trial to compare two treatments. If an optimal design is used, then we hope fewer patients will be required in order to establish the superiority of one of the two treatments and therefore the recruitment and running cost of the trial will be reduced. Furthermore, the most effective treatment can be identified more quickly and reach the population faster, thus improving the quality of life for patients and their carers.

At the moment there is little guidance on how to plan experiments involving possibly censored data. The aim of our work is to fill this gap by incorporating survival models that include censoring in the existing optimal design theory. Novel methods are therefore produced for the construction of optimal designs for many commonly used survival models based on analytical rather than the empirical arguments currently available. Another of our objectives is to provide easily interpreted results that correspond to several scenarios arising in practice. Our results can thus be used directly by practitioners in many relevant situations. This has a potential impact on the planning of survival experiments in the future.

To meet our goals we define a large class of two-parameter nonlinear models that includes some of the most widely used models in practice and which is based on some easily verifiable assumptions. For models in this class we find designs based on two classical optimality criteria, namely D- and c-optimality. As with all optimal experimental designs for nonlinear models, our designs depend on the unknown model parameter values and are referred to as locally optimal designs.

Following this we also investigate the construction of parameter robust designs in a situation where an uncertainty space for the parameter values is provided by the experimenter. These designs can be used if the parameter values are misspecified and so locally optimal designs are not appropriate.

One of the most important survival models is Cox's proportional hazards model for which parameter estimation, and hence the construction of optimal designs, requires a different approach from the one used for parametric models. We therefore study optimal designs for this model separately to further increase the impact of our work on survival experiments.

Another scenario often arising in practice is that of the assumed model to be only an approximation to the true model which leads to the need for model robust designs to be found. We consider a parametric proportional hazards model, the exponential regression model, and construct designs which are robust to small deviations from that model. This is the simplest survival model and is frequently used in survival experiments since an exponential distribution for the times to event can naturally be assumed.

The present thesis is organised as follows. Chapter 2 illustrates the basic concepts of optimal design theory and of Survival Analysis and we also briefly discuss the framework under which these two branches of statistics can be reconciled. The literature on locally optimal, parameter- and model-robust designs for nonlinear models is presented in Chapter 3 together with some methods available for designing survival experiments with censored data. In Chapter 4 we provide analytical characterisations of locally optimal designs based on the D- and c-optimality criteria for the class of nonlinear models we define. Parameter-robust design strategies for the same class of models are discussed in Chapter 5 and used to provide analytical results. Our findings in both Chapters 4 and 5 are illustrated through an application to the exponential regression model under two censoring mechanisms. In Chapter 6 we set up a general framework for the construction of optimal designs for Cox's proportional hazards model and compare the resulting designs with corresponding designs for parametric models. A new class of models in a neighbourhood of the exponential regression model is defined in Chapter 7. This class includes other commonly used survival models. We find designs which are robust to misspecifications of the assumed model within this class. Finally, the conclusions and benefits of our work and possible future directions that can be explored are discussed in Chapter 8.

Chapter 2

Basics

Here we provide a brief introduction to the theory of optimal experimental planning and to the modelling of survival data. For a more detailed presentation of these concepts see, for example, Atkinson, Donev and Tobias (2007) and Collett (2003) respectively. Moreover, we discuss how these two areas of statistical research can come together to produce optimal designs appropriate for survival models which incorporates several censoring mechanisms.

2.1 Optimal experimental planning for parameter estimation

Throughout this project we consider experiments where one is interested in estimating the unknown model parameters. Therefore, we are dealing with an estimation problem and optimal planning of such experiments is concerned with finding the experimental points and the number of subjects that should be assigned to each point so that the parameters are estimated with high precision. This is formulated through an optimal experimental design.

2.1.1 Exact and approximate designs

There are two possible definitions of experimental designs. If m is the number of distinct experimental points in the design then an exact design is defined as

$$\xi_{exact} = \left\{ \begin{matrix} x_1, & \ldots, & x_m \\ r_1, & \ldots, & r_m \end{matrix} \right\},$$

where $0 < r_i \leq n$, $i = 1, \ldots, m$, is the integer number of observations to be taken at the ith experimental point x_i and n is the total number of subjects utilised in the experiment and therefore the total number of observations. Hence $\sum_{i=1}^{m} r_i = n$.

The second definition can be derived using

$$r_i^* = r_i/n, \qquad \sum_i r_i^* = 1.$$

By relaxing the assumption that $nr_i^* = r_i$ must be an integer, we define an approximate design, alternatively known as a continuous design, as

$$\xi = \begin{Bmatrix} x_1, & \dots, & x_m \\ \omega_1, & \dots, & \omega_m \end{Bmatrix}. \tag{2.1}$$

The points x_i, $i = 1, \dots, m$, are called the support points of the design and correspond to the distinct experimental points where observations must be taken and the weights ω_i, $i = 1, \dots, m$ represent the proportion of observations to be taken at the corresponding support point.

The set $\mathcal{X}$ of all possible values for the support points is called the design space. The weights take values $0 \leq \omega_i \leq 1$, $i = 1, \dots, m$ and $\sum_{i=1}^m \omega_i = 1$. Therefore an approximate design ξ is a probability measure on the design space $\mathcal{X}$.

Approximate designs are preferred to exact designs since they are independent of the total number of observations n and their computation avoids the discrete optimisation that is required to find exact designs. However, as pointed out by Atkinson, Donev and Tobias (2007), all designs in practice are exact. Hence if an approximate design is constructed, then for given number of subjects n, the quantity $n\omega_i$ must be rounded to an integer in order for the design to be used. This may result in suboptimal designs for small values of n.

Pukelsheim and Rieder (1992) tackle this problem by introducing a discretisation method called efficient rounding which produces good exact designs for moderate n by rounding the corresponding continuous design. Therefore, in what follows we consider approximate designs of the form (2.1). This provides us with some useful theoretical tools which we discuss in section 2.1.3.

2.1.2 Optimality criteria

The choice of the design to be used is based on optimality criteria which reflect the aim of the experiment to be conducted. In the concept we consider here, estimating the model parameters with high precision means that the asymptotic variance-covariance matrix of the maximum likelihood estimator for the parameters must be minimised. In terms of an approximate design ξ this is equivalent to maximising the information matrix $M(\xi, \boldsymbol{\lambda})$ defined as

$$M(\xi, \boldsymbol{\lambda}) = \int_{\mathcal{X}} I(x, \boldsymbol{\lambda}) \xi(dx) = \sum_{i=1}^m \omega_i I(x_i, \boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda}$ is the parameter vector and $I(x, \boldsymbol{\lambda})$ is the Fisher information matrix. That is, the expectation of the observed information and is given by

$$I(x, \boldsymbol{\lambda}) = E\left[-\frac{\partial^2 \log L(x, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}\right], \tag{2.2}$$

where $\log L(x, \boldsymbol{\lambda})$ is the log-likelihood function at point x for the assumed nonlinear model.

Since we cannot directly optimise a matrix what we actually optimise is a statistically meaningful functional, usually involving only the information matrix, that maps the information matrices onto the real line. This functional, $\Phi\{M(\xi, \boldsymbol{\lambda})\}$, is called the objective function of the criterion and it differs according to the combination of model parameters we are interested in estimating. The aim is then to minimise the objective function with respect to the design ξ to construct the corresponding optimal design.

We note here the dependence of the information and the Fisher information matrices on the vector of unknown model parameters $\boldsymbol{\lambda}$. This is a typical feature that applies only for nonlinear models and not for linear models. Therefore, the optimal designs arising for nonlinear models depend on the values of the parameters and, following Chernoff (1953), they are referred to as locally optimal designs.

Optimality criteria are often symbolised by a letter of the alphabet and hence are sometimes called alphabetical optimality criteria (Atkinson, Donev and Tobias (2007)). Two of the most popular ones which we consider throughout this project are D- and c-optimality. These are used when one is interested in estimating all the model parameters or a linear combination of them respectively and are explicitly defined in Chapter 4.

2.1.3 General equivalence and Caratheodory's theorems

The general equivalence theorem is a very useful tool for the characterisation and checking of optimal designs. This does not hold in general for exact designs but only for approximate designs. If ξ^* is the optimal design, the general equivalence theorem states that the following three statements are equivalent (Atkinson, Donev and Tobias (2007)).

(i) The design ξ^* minimises $\Phi\{M(\xi, \boldsymbol{\lambda})\}$.

(ii) Let $\phi(x, \xi, \boldsymbol{\lambda})$ be the derivative of Φ in the direction $\tilde{\xi}$ given by

$$\phi(x, \xi, \boldsymbol{\lambda}) = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\left[\Phi\left\{(1-\epsilon)M(\xi, \boldsymbol{\lambda}) + \epsilon M(\tilde{\xi}, \boldsymbol{\lambda})\right\} - \Phi\left\{M(\xi, \boldsymbol{\lambda})\right\}\right],$$

where $\tilde{\xi}$ is the design putting all the observations at point x. The design ξ^* maximises the minimum over $\mathcal{X}$ of $\phi(x, \xi, \boldsymbol{\lambda})$.

(iii) The minimum over $\mathcal{X}$ of $\phi(x, \xi^*, \boldsymbol{\lambda})$ is equal to zero and this minimum is achieved at the support points of the design.

Statement (iii) results in the further statement

(iv) The minimum over $\mathcal{X}$ of $\phi(x, \xi, \boldsymbol{\lambda})$ is strictly negative for any suboptimal design ξ.

The special cases of the general equivalence theorem for D- and c-optimality respectively, are provided in Chapter 4. An upper bound for the number of support points of the optimal design can be obtained due to the additive nature of the information matrix (see, for example, Atkinson, Donev and Tobias (2007) or Silvey (1980)). This is Caratheodory's theorem, which states that the optimal design must be supported at most at $p(p+1)/2$ points where p is the total number of model parameters. When a design has exactly p support points it is said to be minimally supported. In Chapter 4, we use the general equivalence theorem to establish that the D-optimal designs for the class of models considered, are minimally supported.

2.1.4 Bayesian designs

The Bayesian approach for the construction of optimal designs takes into account any prior information available for the parameter vector $\boldsymbol{\lambda}$. Let $\boldsymbol{y}$ be the vector of n observations indicating the data and $\pi(\boldsymbol{\lambda})$ the prior distribution for the parameter vector. The Bayesian optimal design maximises

$$U(\xi) = \int \log\{\pi(\boldsymbol{\lambda}|\boldsymbol{y}, \xi)\}\, \pi(\boldsymbol{y}, \boldsymbol{\lambda}|\xi)\, d\boldsymbol{\lambda}\, d\boldsymbol{y},$$

which is the expected Shannon information of the posterior distribution of $\boldsymbol{\lambda}$. A detailed presentation of optimal Bayesian designs is given in Chaloner and Verdinelli (1995).

In the context of this project we do not use Bayesian optimal designs as there are several references in the literature of other designs that are simpler to find and equally efficient for parameter estimation as Bayesian designs (see, for example, Dror and Steinberg (2006) or Biedermann and Woods (2011)). In particular, we will focus on the construction of standardised maximin optimal and cluster designs when a set of parameter values is provided by the experimenter with no preference for specific values and therefore a Bayesian framework is unnecessary (see Chapter 5).

2.2 Modelling time to event data

As mentioned in the introduction, the response variable arising in survival experiments is the time until the event of interest occurs. However, the event of interest may not be observed for some of the subjects utilised in the experiment, leading to censored survival data. In what follows we present the two censoring mechanisms we shall consider and two classes of survival models widely used for fitting time to event data.

2.2.1 Censoring mechanisms

The most common form of censoring arising in practice is right-censoring. In this case the time until the occurrence of the event of interest is above a certain value called the censoring time, but it is unknown by how much (Collett (2003)). Therefore, for each subject, if the event of interest is not observed by the censoring time for that subject, its observation is said to be right-censored.

Let us consider the case of a clinical trial with death as the event of interest, where a patient drops out from the trial due to worsening of their health. The time of death of this person is therefore censored since the death is not observed. However, it is clear in this case that the death time and censoring time are associated. This is an example of informative censoring and the methods for analysing survival data subject to this mechanism are different from the ones discussed here. In particular, we assume non-informative censoring, that is, for each subject the time until the occurrence of the event of interest is statistically independent of its censoring time.

There are several mechanisms that result in right-censored data. The two most usual ones that we consider throughout this project are Type-I censoring where the censoring time is fixed and common for all the subjects in the experiment and random censoring in which case the censoring time is possibly different for each subject.

2.2.2 Survivor and hazard functions

Time to event data are summarised mainly using the survivor and hazard functions (see Collett (2003)). Let T be the random variable indicating the time to event with t being its observed value. The survivor function evaluated at point t, $S(t)$, is defined as the probability that the event of interest will occur at some time greater than or equal to t. That is

$$S(t) = P(T \geq t) = 1 - F(t), \qquad t > 0,$$

where $F(t)$ is the distribution function of the random variable T.

The hazard function $h(t)$ expresses the risk of the event occurring at any time t after the commencement of the experiment. It is defined as the event rate at time t conditional on the event occurring at or before t. Hence

$$h(t) = \lim_{\delta t} \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t}$$

$$= \lim_{\delta t} \frac{F(t+\delta t) - F(t)}{S(t)\delta t} = \frac{f(t)}{S(t)}, \qquad t > 0,$$

where $f(t)$ is the probability density function of T.

In survival experiments one is interested in exploring how the risk of occurrence of the event of interest changes with respect to various factors. Therefore, the hazard function is modelled directly and these factors are referred to as the explanatory variables or the covariates of the model.

2.2.3 Proportional hazards models

One of the most popular classes of survival models is that of proportional hazards models. The main assumption governing all survival models included in this class is the proportional hazards assumption which states that the explanatory variables involved in the model have a multiplicative effect on the hazard. In other words at any time $t(> 0)$ the hazard function of a subject with a certain vector of covariate values is proportional to the hazard function of another subject and therefore their hazard ratio is constant over time. The general form of proportional hazards models is given by

$$h(t) = h_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{x}}, \qquad t > 0, \tag{2.3}$$

where $\boldsymbol{x}$ is the vector of explanatory variable values, $\boldsymbol{\beta}$ is the corresponding covariate coefficients vector and $h_0(t)$ is the hazard function for a subject with $\boldsymbol{x} = 0$ and is called the baseline hazard function.

When a specific form is assumed for the baseline hazard function the resulting models are referred to as parametric proportional hazards models. The most frequently used distributions for the specification of such models are the exponential, Weibull and Gompertz distributions. If no particular form for the baseline hazard function is specified then model (2.3) is Cox's proportional hazards model which is often preferred to parametric proportional hazards models because fewer assumptions are required.

When introducing this model, Cox (1972) showed that inferences on the β-coefficients can still be done independently of $h_0(t)$ and based only on the order of occurrence of the events corresponding to the various subjects. The main assumption he uses, apart from that of proportional hazards, is that the baseline hazard and so the hazard function is zero in time intervals in which the event of interest has not occurred for any of the subjects. Therefore, these intervals provide no information about the β-coefficients.

Let $t_{(1)} < \ldots < t_{(\tilde{n})}$ be the ordered distinct event times independent of one another and the rest of the available data, $n - \tilde{n}$, are right-censored observations. Let us also assume that each event time $t_{(j)}$, $j = 1, \ldots, \tilde{n}$ corresponds to exactly one subject in the experiment, that is, there are no ties in the data. The probability of the event occurring at some time $t_{(j)}$ for a subject with covariate values vector $\boldsymbol{x}_{(j)}$ conditional on $t_{(j)}$ being one of the distinct event times is given by

$$\begin{aligned}
&P(\text{event occurs at } t_{(j)} \text{ for subject with covariates } \boldsymbol{x}_{(j)} | \text{one event at } t_{(j)}) \\
&= \frac{P(\text{event occurs at } t_{(j)} \text{ for subject with covariates } \boldsymbol{x}_{(j)})}{P(\text{one event at } t_{(j)})} \\
&= \frac{P(\text{event occurs at } t_{(j)} \text{ for subject with covariates } \boldsymbol{x}_{(j)})}{\sum_{l \in R(t_{(j)})} P(\text{event occurs for subject } l \text{ at } t_{(j)})},
\end{aligned}$$

since the event times are independent. The set $R(t_{(j)})$, referred to as the risk set, denotes the set of all subjects that are at risk at time $t_{(j)}$; that is, the subjects for which neither the event of interest nor censoring has occurred at a time just prior to $t_{(j)}$. Furthermore, the above expression is equal to

$$\begin{aligned}
\lim_{\delta t \to 0} & \frac{P(\text{event occurs at } (t_{(j)}, t_{(j)} + \delta t) \text{ for subject with covariates } \boldsymbol{x}_{(j)})/\delta t}{\sum_{l \in R(t_{(j)})} P(\text{event occurs for subject } l \text{ at } (t_{(j)}, t_{(j)} + \delta t))/\delta t} \\
&= \frac{h_j(t_{(j)})}{\sum_{l \in R(t_{(j)})} h_l(t_{(j)})},
\end{aligned}$$

using the definition of the hazard function given in section 2.2.2. Here $h_j(t_{(j)})$ and $h_l(t_{(j)})$ are the hazard functions at time $t_{(j)}$ for a subject with covariate values vector $\boldsymbol{x}_{(j)}$ and $\boldsymbol{x}_{(l)}$ respectively. Now using equation (2.3) and taking the product of these conditional probabilities over the distinct event times, we obtain

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{\tilde{n}} \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_{(j)}}}{\sum_{l \in R(t_{(j)})} e^{\boldsymbol{\beta}^T \boldsymbol{x}_l}}. \tag{2.4}$$

This is the likelihood function to be used for the estimation of the β-coefficients and is called the partial likelihood function.

Kalbfleisch and Prentice (2002) derive the exact form of the partial likelihood function in the case of ties in the data which, however, is extremely complicated. Some approximations of the partial likelihood function given in (2.4) are suggested by Cox (1972), Breslow (1974) and Efron (1977) which are easier to compute.

2.2.4 Accelerated failure time models

An alternative to the proportional hazards models is the class of accelerated failure time models (see, for example, Collett (2003)) which are specified by the hazard function

$$h(t) = h_0(t/e^{\boldsymbol{\beta}^T \boldsymbol{x}})e^{-\boldsymbol{\beta}^T \boldsymbol{x}}, \qquad t > 0,$$

or equivalently by the survivor function

$$S(t) = S_0(t/e^{\boldsymbol{\beta}^T\boldsymbol{x}}), \qquad t > 0.$$

The quantity $e^{-\boldsymbol{\beta}^T\boldsymbol{x}}$ is called the acceleration factor and $h_0(t)$, $S_0(t)$ are the baseline hazard and survivor functions respectively. As for the proportional hazards models, these can be regarded as the hazard and survivor functions for a subject for which all covariate values are equal to zero.

Under an accelerated failure time model the explanatory variables have a multiplicative effect on the survivor times but the assumption of constant hazard ratio over time is not satisfied.

Accelerated failure time models are met, for example, in survival experiments where the subjects utilised are put under extreme conditions so that the event of interest will occur sooner than under normal circumstances. Such models are not commonly used for data arising in clinical trials but are frequently used in industrial applications.

2.3 Optimal designs for survival models

Let us now consider that an experimental design is required before the commencement of a survival experiment with predetermined total duration c. The design must be optimal in terms of estimating the unknown parameters of the assumed survival model. Throughout this project we consider two cases: a binary design space, that is $\mathcal{X} = \{0, 1\}$, corresponding to a covariate indicating, for example, two different treatments; and the case of a continuous design space $\mathcal{X} = [u, v]$ corresponding to a covariate representing, for example, the doses of a drug.

Also let $y_1, \ldots, y_n$ be the possibly right-censored data that will arise from the survival experiment utilising a total number of n subjects. These are the observed values of the random variables $Y_j = \min\{T_j, C_j\}$, $j = 1, \ldots, n$, where T_j, $j = 1, \ldots, n$, indicate the time until the occurrence of the event of interest for each subject and are distributed according to the assumed model. The variable C_j represents the censoring time corresponding to the jth subject (Collett (2003)). Hence if the event of interest has not occurred for the jth subject before its corresponding censoring time value c_j, then the observation is considered to be right-censored and is equal to c_j. This is formulated using an indicator variable δ_j that is equal to unity if the observation is a distinct event time and zero if it is right-censored. That is

$$\delta_j = \begin{cases} 1, & \text{if } Y_j = T_j \\ 0, & \text{if } Y_j = C_j \end{cases}.$$

Unlike data arising in the absence of censoring where the likelihood function is given as the product of the probability density function evaluated at each data point, the likelihood function for censored data is

$$L(\boldsymbol{\lambda}) = \prod_{j=1}^{n} \{f(y_j)\}^{\delta_j} \{S(y_j)\}^{1-\delta_j}, \tag{2.5}$$

where $\boldsymbol{\lambda}$ is the vector of the unknown model parameters, $f(y_j)$ is the probability density function of the assumed distribution for the times to event and $S(y_j)$ the corresponding survivor function.

Therefore, the Fisher information matrix defined in (2.2) will differ for censored data, not only because of the different form of the likelihood function described above but it will also depend on the censoring mechanism giving rise to such data through the expectations involved in (2.2). Hence the assumed type of right-censoring affects the information matrix and thus the resulting optimal design.

2.3.1 Type-I censoring

As mentioned in section 2.2.1, under this mechanism the censoring time must be the same for all the subjects in the experiment (Collett (2003)). In the scenario we consider here this corresponds to the case of all the subjects being recruited to the experiment at the same time and so the common censoring time will be equal to c, that is, the duration of the experiment.

Therefore, the random variable Y_j will follow the assumed distribution for the times to event until time c at which point if the event of interest has not occurred, $Y_j = c$. The probability of this happening is equal to the probability that the corresponding time to event variable T_j will be greater than or equal to c. That is, $P(T_j \geq c) = S(c)$ from the definition of the survivor function given in section 2.2.2. Hence

$$E(Y_j) = \int_0^c y_j f(y_i)\, dy_j + cP(Y_j = c) = \int_0^c y_j f(y_i)\, dy_j + cS(c).$$

2.3.2 Random censoring

For the case of random censoring we consider the situation where the subjects enter the experiment at random times Z_j, $j = 1, \ldots, n$ uniformly distributed in the time interval $[0, c]$. Hence the censoring times $C_j = c - Z_j$, $j = 1, \ldots, n$ will possibly be different for each subject and $C_j \sim U[0, c]$, $j = 1, \ldots, n$ with probability density function $f_C(c_j) = 1/c$.

The probability density function function that must be used for the likelihood function given in (2.5) is now $f(y_j|c_j) = \frac{1}{c} f(y_j)$ since we assume non-informative censoring, that is, the event times are independent of the censoring times. Furthermore,

$$E(Y_j) = E\left(E(Y_j|C_j = c_j)\right) = \int_0^c E(Y_j|C_j = c_j)\, dc_j,$$

where

$$E(Y_j|C_j = c_j) = \int_0^{c_j} y_j f(y_i)\, dy_j + c_j S(c_j).$$

Chapter 3

Literature review

In this chapter we provide a review on the available literature relevant to this project. The biggest part of this literature is concerned with the optimal planning of experiments involving nonlinear models with complete data, that is, data that are not subject to censoring. We first discuss the construction of locally optimal designs based on criteria that include D- and c-optimality which are of primary interest in this thesis, for several classes of nonlinear models. Various techniques for overcoming the parameter dependence of the locally optimal designs are then presented. These can be used for the construction of parameter robust designs when an uncertainty space can be specified for the model parameter values. Finally, we present model robust designs that are appropriate for use when the assumed nonlinear model only holds approximately.

Little research has been done, however, on designing experiments using the optimal design theory for possibly censored data. The literature is mainly focused on experiments involving accelerated failure time models and particularly on finding designs which are robust to misspecifications in the underlying distribution of these models. Fewer authors study the construction of optimal designs using the classical optimality criteria for proportional hazards models. We review the limited number of available papers that consider the exponential regression model in its proportional hazards parametrisation and papers that use the general proportional hazards model. In the latter case the designs are constructed using the partial likelihood function.

3.1 Locally optimal designs

One of the most important general results available for locally optimal designs is given in Pukelsheim and Torsney (1991) who derive explicit formulas for the optimal weights of designs constructed based on a broad class of criteria called Φ_p-optimality criteria. This class is introduced by Kiefer (1974) and includes the popular D-optimality criterion. Although Pukelsheim and Torsney (1991) consider the classical linear regression model many authors, including ourselves, use their result to find the corresponding formulas for the Φ_p-optimal weights when the models involved are nonlinear.

Ford, Torsney and Wu (1992) consider the construction of locally D- and c-optimal designs for nonlinear models where the distribution of their response variable is a member of the exponential family. They propose the transformation of the design space which leads to a simpler design problem in a canonical form. Using Elfving (1952) and Sibson (1972) geometrical characterisations they find locally D- and c-optimal designs for various two-parameter regression models involving one explanatory variable.

These results on locally D-optimal designs are then extended by Sitter and Torsney (1995a) who consider generalised linear models involving several explanatory variables. They show that, based on the geometry of the transformed design space, the design problem can be reduced in terms of complexity to that of only one explanatory variable. The transformed design space proposed by Ford, Torsney and Wu (1992) is also used by Sitter and Torsney (1995b), who focus on binary response models and construct locally D- and c-optimal designs for the case of two design variables by using geometrical arguments.

Sebastiani and Settimi (1997) consider the two-parameter logistic regression model and prove that the two-point design suggested by Ford, Torsney and Wu (1992) is D-optimal for this model. The cases of a design space bounded at one end and at both ends are investigated separately. Moreover, using approximations they find designs that do not require exact knowledge of the model parameter values, unlike the locally optimal designs, and show that even though these designs are not optimal, they are efficient alternatives to the locally D-optimal designs.

A new geometrical interpretation of Φ_p-optimal designs is provided by Biedermann, Dette and Zhu (2006) for two-parameter regression models based on the idea of the minimum confidence ellipsoid used for the classical D-optimality criterion. This result also offers some intuition on both the position and the number of support points of the optimal designs. They apply this method to binary response models for a wide class of link functions and construct Φ_p-optimal designs for both a bounded and an unbounded design space. Finally, they show that the Φ_p-optimal designs are minimally supported if a condition involving the link function of the corresponding model is satisfied. We have used a modification of the idea used in the proof of this latter result to show that the locally D-optimal designs for the class of nonlinear models we consider are minimally supported (see Lemma 1 in section 4.2.3).

Russell et al. (2009) focus on Poisson regression models involving one or more explanatory variables and having a log-linear link function. Using the Ford, Torsney and Wu (1992) canonical form for the design problem, they provide a theoretical result on locally D-optimal designs for such models. In section 4.2.3 we show that our analytical characterisation of locally D-optimal designs generalises their result in the case of one explanatory variable.

In recent years, producing general results for a class of models has become popular in the optimal design literature. Hedayat, Zhong and Nie (2004) define a class of two-parameter nonlinear models based on some assumptions on the Fisher information matrix and show that for these models the locally D-optimal designs are minimally supported. They also provide analytical and geometrical methods for the construction of designs which are efficient for parameter estimation, although not optimal. However, their assumptions on the Fisher information matrix are not generally satisfied. For example, these results are not applicable to the exponential-based proportional hazards model for censored data which we consider in section 4.4.

An even more general class of nonlinear models is considered by Yang and Stufken (2009) who find optimal designs using Loewner optimality. They obtain a series of excellent results that show, depending on some conditions, for each given design there is always another design from a simple class which is better in the Loewner sense and hence it is also better under commonly used criteria such as Φ_p-optimality criteria. These results are then generalised to nonlinear models with more than two unknown model parameters by Yang (2010), Dette and Melas (2011), Yang and Stufken (2012) and Dette and Schorning (2013). However, the conditions necessary for the derivation of the results in these papers can be difficult to verify even using symbolic computational software.

3.2 Parameter-robust designs

As mentioned above, optimal experimental designs for nonlinear models depend on the true values of the model parameters. In many practical situations an uncertainty space for the parameter values can be specified. Therefore, many authors consider the construction of designs that are robust to misspecifications of the parameter values and hence perform well across the specified uncertainty space. Such design strategies are the construction of maximin, Bayesian and cluster designs which are discussed separately in the following sections.

3.2.1 Maximin designs

A maximin design maximises the corresponding optimality criterion function with respect to the design for those parameter values in the uncertainty space for which the function is minimised. Haines (1995) considers nonlinear models that involve only one unknown parameter and presents a geometrical method for constructing maximin designs when a range of parameter values is specified. The extension of this approach to models with more than one unknown parameter is not, however, straightforward.

Dette (1997) introduces a class of standardised maximin optimality criteria that are invariant under linear transformations of the design space. The designs constructed using these criteria maximise the minimum of the ratio of the criterion function evaluated at the locally optimal design over the criterion function for an arbitrary design across the uncertainty parameter space. Due to the useful invariance property which, for the classical criteria only holds for D-optimality, this standardised approach has become very popular.

Minimax designs, introduced by Elfving (1959), that minimise the maximum variance are considered by Dette and Sahm (1998) for binary response models. They compare the designs found based on this classical criterion with designs for the standardised version of the criterion following Dette (1997), and find that the former designs should not be preferred since the number of their support points is found to be less than the number of model parameters. Imhof and Wong (2000) propose a general graphical method for finding maximin designs which can be used, however, only to determine candidate designs since they provide no theoretical characterisations of these designs. Some analytical results on maximin designs for various heteroscedastic polynomial models are given in Imhof (2001).

Dette and Biedermann (2003) consider the construction of standardised maximin D-optimal designs for the Michaelis-Mentel model which is often encountered in biology studies. Given a range of reasonable parameter values and following Dette (1997), the designs are found by maximising the minimum efficiencies over the range of parameter values. Closed form expressions for the locally D-optimal designs are first derived and then Dette and Biedermann (2003) provide analytical characterisations of the standardised maximin D-optimal designs supported at exactly two points. Moreover, they assess the performance of their designs through a real data example and illustrate that these designs are highly efficient in the case of the parameter values being misspecified.

3.2.2 Bayesian designs

The Bayesian analogues of alphabetical criteria and other design criteria within the Bayesian framework for nonlinear models are presented in Chaloner and Verdinelli (1995). They show that number of support points of the Bayesian optimal designs depend on the prior distribution assumed for the unknown model parameters. Bayesian optimal designs for nonlinear models are also constructed using a geometrical approach in Haines (1995) when only one unknown parameter is involved.

Throughout this project we assume that there is no preference for specific values in the given parameter space and therefore a Bayesian approach requiring the specification of a prior distribution for the parameters is not necessary. Moreover, as will be discussed in the next section, Bayesian optimal designs are found to have similar performance as other parameter robust designs available in the literature, which are easier to find.

3.2.3 Cluster designs

The cluster design strategy was introduced by Dror and Steinberg (2006) for generalised linear models involving several explanatory variables. Considering D-optimality, they first compute the locally optimal designs corresponding to several parameter values drawn from the specified uncertainty space which they then combine into position vectors and apply a K-means clustering procedure to obtain the resulting cluster design. The robustness of their designs is assessed through a simulation study for possibly misspecified parameter values, linear predictors and link functions. Among other comparisons, Dror and Steinberg (2006) compare cluster designs to Bayesian designs and find that they perform similarly. Also taking into account that the former designs are more easily computed, they conclude that cluster designs are good alternatives to the more sophisticated method of the construction of Bayesian optimal designs.

The method proposed by Dror and Steinberg (2006) is used by Russell et al. (2009) for the construction of cluster designs for multivariate Poisson regression models. More recently, Biedermann and Woods (2011) modify this algorithm so that the weights of the cluster designs are allowed to be unequal. Through an application to second harmonic generation experiments, Biedermann and Woods (2011) illustrate that cluster designs are effective and more easily computed alternatives to Bayesian optimal designs.

3.3 Model-robust designs

One of the first references about criteria that can be used for the construction of optimal designs when the assumed model is incorrect is Wiens (1992). He considers the problem of precise estimation of model parameters when the assumed linear regression model holds only approximately. The optimality criteria he proposes correspond to various classical alphabetical criteria but are based on the mean squared error matrix. Minimax designs are constructed such that they minimise the criteria functions for the worst possible deviation from the linear regression model. In Chapter 7 we extend this method to the exponential regression model.

Sinha and Wiens (2002) consider the construction of sequential designs for approximately specified nonlinear regression models. That is, given a prior estimate for the model parameter values, the design takes into account any information obtained during the course of the experiment. The vector of parameter estimates is then updated and this procedure continues until the desired estimation accuracy is achieved or until the available resources are exhausted. We are concerned with designing experiments prior to their commencement and in this context the construction of such sequential designs is not realistic.

Possibly misspecified nonlinear regression models are also studied by Wiens and Xu (2008a). They find minimax designs for the extrapolation of the response to a point outside the design space. Extrapolation problems are usually considered in accelerated failure time tests where an extrapolation to lower values of the explanatory variables is required. The results of Wiens and Xu (2008a) are extended by Wiens and Xu (2008b) for both extrapolation and prediction problems, the latter corresponding to the case in which one is interested in the response. Xu (2009a) studies the construction of maximin designs for approximate exponential regression models. He considers both the cases of homoscedacity and heteroscedacity for prediction problems. However, neither of the papers discussed here, refers to the problem of estimating the model parameters which we are interested in.

Woods et al. (2006) consider generalised linear models and propose a method for the construction of exact designs based on what they call compromise design selection criteria. The resulting exact designs are robust to misspecifications of the link function, the linear predictor and of the model parameter values. A comparison of compromise designs and cluster designs is performed by Dror and Steinberg (2006) who show that cluster designs constitute a better robust design strategy due to their simplicity and the short computational time of their construction.

3.4 Optimal designs for survival models

We now present the available literature on the construction of optimal designs for the most popular survival models fitted to possibly censored data. Accelerated failure time models are considered first followed by the exponential regression model in its proportional hazards parametrisation and finally we discuss the construction of optimal designs for Cox's proportional hazards model.

3.4.1 Accelerated failure time models

Pascual and Montepiedra (2003) define a criterion for the construction of designs which are robust to model uncertainty when interest is in estimating quantiles. An equivalence theorem is also presented that can be used to check the optimality of candidate designs based on this criterion. They illustrate their results through an application to a practical accelerated time test for which uncertainty lies in using the Weibull or the log-normal based accelerated failure time model.

A Bayesian approach is considered by Zhang and Meeker (2006) for censored data arising in an accelerated failure time framework. In their application they use a Weibull distribution and Type-I censoring. Wu, Lin and Chen (2006) assume an exponential distribution for the failure times and a step-stress experiments with progressive Type-I censoring mechanism. That is, the subjects are studied in discrete rather than in continuous time intervals and if failure does not occur in that interval the stress level, that is, the value of the explanatory variable, is increased. They consider maximum likelihood estimation using both a minimum variance criterion and D-optimality. However, their designs are not overall optimal but provide only an initial guideline as to how to plan such experiments.

Generalised linear models with normal underlying distribution for censored data are considered by Xu (2009b). He defines criteria for the construction of designs which are robust to misspecifications in the regression function for prediction and extrapolation problems. Although we focus on estimation problems, in Chapter 7 we follow the method used in Xu (2009b) for the derivation of the asymptotic distribution of the maximum likelihood estimator for the parameter vector.

McGree and Eccleston (2010) propose the use of compound optimality criteria that ensure precise estimation of the model parameters and at the same time minimise the time to failure which therefore reduces the experimentation time. They mainly focus on the Weibull-based accelerated failure time model with one parameter and they construct optimal designs that are based on the compound optimality criteria they propose and the compromise criterion introduced in Woods et al. (2006).

The design problem of allocating patients to two treatments in two stages is considered by Bandyopadhyay, Biswas and Bhattacharya (2010). Under this scenario, a few patients are first randomised to the treatments and the information accumulated and the patients' prognostic factors are then taken into account for the allocation of future patients.

3.4.2 Exponential regression model

Locally D-optimal designs for one and two-parameter exponential regression models are studied by Becker, McDonald and Khoo (1989). They investigate the effect of censoring on the locally optimal designs using geometrical arguments and empirical values of the parameters. For the two-parameter model they also discuss the effect of censoring for different shapes of the design space. However, their statements are supported only empirically and the uniqueness of their designs is not proved in general.

López-Fidalgo, Rivas-López and Del Campo (2009) consider a two-parameter exponential regression model which due to its parametrisation requires some constraints on the parameters. They propose an algorithm for the construction of D-optimal designs that depends on the arrival times of the subjects utilised in the experiment. Therefore, whenever a subject enters the experiment a new design has to be found. They study the simple case of allocating subjects, for example, to two treatments, and apply their algorithm by assuming uniform discrete and continuous distributions for the arrival times separately. In the framework considered in this project the experiment is designed beforehand and so designs conditional on arrival times are not appropriate.

3.4.3 Cox's proportional hazards model

The Cox proportional hazards model was introduced by Cox (1972) and since the baseline hazard is of arbitrary form estimation of the model parameters must be done using the partial likelihood function. Cox (1975) shows that the partial likelihood estimators satisfy the same type of asymptotic properties as those for the parameter estimators found using the full likelihood approach. The efficiency of partial likelihood estimation is assessed by Efron (1977). He shows that the Fisher information matrices for the full and partial likelihood methods coincide except for an extra term in the Fisher information for the full likelihood, which, however, will usually be small in practice. Therefore inferences based on the partial likelihood function are similar to the ordinary likelihood approach. Moreover, Andersen and Gill (1982) derive an explicit characterisation of the asymptotic distribution of the partial likelihood estimators for the parameters using a counting process framework for Cox's model.

Despite the results discussed above on inferences based on the partial likelihood approach, little research has been done on how to design experiments for censored data when Cox's proportional hazards model is assumed. To the best of our knowledge the available literature is restricted to two papers.

Kalish and Harrington (1988) consider the problem of allocating patients to two treatments and find optimal designs by minimising the asymptotic variance of the partial likelihood estimate using the results by Andersen and Gill (1982). Assuming a constant baseline hazard, which corresponds to the exponential based proportional hazards model, they find the balanced design that allocates equal proportions of patients to the two treatments to be very efficient for both full and partial likelihood estimation. Their most important result is, however, that the form of the baseline hazard function does not affect the optimal choice of design when the data are subject to Type-I censoring. In Chapter 6 we extend this result to the case of a continuous design space.

An approximation of the partial likelihood Fisher information matrix is proposed by López-Fidalgo and Rivas-López (2012). They use this approximate matrix to construct optimal designs on a binary design space for an exponential regression model and then compare these designs to the ones constructed using the full likelihood approach. However, we found the quality of their approximation to be unsatisfactory when we compared their result on the information matrix to the asymptotic variance matrix provided by Andersen and Gill (1982).

Chapter 4

Locally optimal designs

In this chapter we construct optimal designs for a general class of nonlinear models involving one explanatory variable and therefore having only two parameters. This class is identified by the form and some extra conditions on the information matrix, thus the design problem is solved in more generality. We focus on two-parameter nonlinear models because the experiments of our interest are usually conducted in order to evaluate a particular method or treatment and therefore, the models used involve only one design variable. Moreover, these are the most frequently considered nonlinear models in the literature and our results extend some of the available ones to include models subject to several censoring mechanisms. This enables us to identify how censoring affects the optimal choice of design.

For the construction of the optimal designs, we consider the widely used criteria of D- and c-optimality which correspond to the cases where we are interested in estimating both and one of the two model parameters respectively. As mentioned in the introduction, optimal designs for nonlinear models depend on the unknown model parameters. Hence, they cannot be evaluated in practice and are referred to as locally optimal designs. However, analytical characterisations of locally optimal designs such as the ones we provide here are very important since, as Ford, Torsney and Wu (1992) noted, these designs are vital for the construction of sequential as well as non-sequential designs.

We first introduce the class of nonlinear models to be considered and present examples of models included in this class. The construction of locally D- and c-optimal designs is then examined separately and the cases of a binary and of a continuous design space are explored. We provide analytical characterisations of locally optimal designs for models within the class, thus reducing the numerical effort for design search substantially. Finally, we apply our results to the exponential regression model in its proportional hazards parametrisation and discuss how the optimal designs change in the presence of Type-I and random censoring.

4.1 Class of models

The class of two-parameter nonlinear models is now defined based on the form of and some extra assumptions on the information matrix for an arbitrary design ξ. We briefly discuss models studied in the literature that share the same form of information matrix but satisfy different conditions than the ones we define and give examples of models that are are included in the class we consider.

4.1.1 Information matrix assumptions

Let ξ be an approximate design with support points x_i, $i = 1, \ldots, m$ taking values in the design space $\mathcal{X}$ and corresponding weights ω_i, where $0 < \omega_i \leq 1$, $i = 1, \ldots, m$ and $\sum_{i=1}^{m} \omega_i = 1$. We consider two-parameter nonlinear models with information matrix of the form

$$M(\xi, \alpha, \beta) = \sum_{i=1}^{m} \omega_i I(x_i, \alpha, \beta) = \sum_{i=1}^{m} \omega_i Q(\theta_i) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}, \tag{4.1}$$

where $I(x_i, \alpha, \beta)$ is the Fisher information matrix at the point x_i, α and β are the unknown model parameters and $\theta_i = \alpha + \beta x_i$.

Many authors have studied models with the same type of information matrix as the one defined in (4.1). Ford, Torsney and Wu (1992) focus on generalised linear models where the response variable is distributed as a member of the exponential family. They consider three different forms for their corresponding Q-function but none of these forms is applicable for proportional hazards models subject to censoring (see, for example, section 4.4.3). Two-parameter nonlinear models with information matrix of the form (4.1) are also studied by Hedayat, Zhong and Nie (2004). Their key assumptions are that $Q(\theta)$ has exactly one stationary point and $\lim_{\theta\to\infty} Q(\theta)\theta^2$ is bounded. However, for many relevant situations $Q(\theta)$ is strictly increasing and $\lim_{\theta\to\infty} Q(\theta)\theta^2$ is unbounded (see, for example, section 4.4.3). A more general class of models and Loewner optimality are considered by Yang and Stufken (2009) who obtained excellent results, showing that under some conditions, for each given design there is always a design from a simple class which is better in the Loewner sense. Depending on the model, however, the conditions can be difficult to verify, even with symbolic computational software.

We focus on models with information matrix of the form (4.1) which also satisfy conditions (a)-(d) and (d1) given below. Following Ford, Torsney and Wu (1992), for the definition of these conditions we consider the transformed design space $\Theta = \alpha + \beta\mathcal{X}$, where $\beta \neq 0$. The parameter dependence of the design problem thus enters only via the transformed design space. We note that for $\beta = 0$ we have $Q(\theta) = Q(\alpha)$ which corresponds to the trivial case of a linear model. The conditions are given for $\theta \in \mathbb{R}$, so they are valid for all possible ranges of Θ.

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

(b) The function $Q(\theta)$ is strictly increasing on $\mathbb{R}$.

(c) The second derivative $g_1''(\theta)$ of the function $g_1(\theta) = 2/Q(\theta)$ is an injective function.

(d) For any $s \in \mathbb{R}$, the function $g_2(\theta) = Q(\theta)(s-\theta)^2$ satisfies $g_2'(\theta) = 0$ for exactly two values of $\theta \in (-\infty, s]$.

For the case of c-optimality we require the extra condition

(d1) : The function $\log Q(\theta)$ is concave for $\theta \in \mathbb{R}$,

which implies condition (d) given that (a) and (b) are satisfied, the proof of which is given in section A.1.1 of Appendix A.

Our aim throughout this project is to produce results that can be easily interpreted and therefore implemented directly by practitioners, particularly working in experiments involving survival models. We thus define easily verifiable conditions which however are satisfied by many models widely used in practice also in the presence of censoring.

4.1.2 Examples

The generalised linear model with response variable following a Poisson distribution is included in the class of models considered. For example, if we assume a Pois($e^{\alpha+\beta x}$) distribution the corresponding log-likelihood function results in the Fisher information matrix at point x given by

$$I(x, \alpha, \beta) = e^{\alpha+\beta x} \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix},$$

which yields (4.1) with $Q(\theta) = e^\theta$. This function is positive for all $\theta \in \mathbb{R}$ with $Q'(\theta) = Q''(\theta) = e^\theta$ and so conditions (a) and (b) are satisfied. Moreover, it is easy to see that the function $g_1''(\theta) = 2/e^\theta$ defined in condition (c) is decreasing with θ and hence injective. Finally, condition (d1) and therefore condition (d), holds as $(\log Q(\theta))'' = 0$. This form of Poisson regression with rate dependent on the explanatory variable in a log-linear manner is an example of a model resulting in the second form for the Q-function considered by Ford, Torsney and Wu (1992). It is also studied by Russell et al. (2009) for possibly more than one independent variables.

Further examples of generalised linear models satisfying our assumptions on the information matrix are the ones with response variable following a Gamma($\gamma, (k_1 + e^{\alpha+\beta x})^{k_2}$) or an Inv-Gamma($\gamma, (k_1 + e^{\alpha+\beta x})^{k_2}$) distribution, where $\gamma (> 0)$ is the shape parameter of the distribution, $k_1 > 0$ and $k_2 \neq 0$ are constants and all γ, k_1 and k_2 are assumed known (see section A.1.2 in appendix A for a proof).

The class under our consideration also includes any parametric proportional hazards model with hazard function of the form

$$h(t, x) = e^{\alpha} r(t) e^{\beta x}, \quad r(t), t > 0 \tag{4.2}$$

and response variable subject to Type-I censoring. When the response variable is subject to random censoring we also require the extra condition of the function $\int_0^{c_j} r(s)\,ds$ being log-concave in c_j on $\mathbb{R}^+$ (see sections A.1.3 and A.1.4 respectively of Appendix A for proofs). The expression $e^{\alpha} r(t)$ is the assumed baseline hazard function and any parameters involved in $r(t)$, such as the shape parameter γ of the assumed distribution, are considered known. Examples of such models are the parametric proportional hazards model based on the Exponential, Weibull and Gompertz distribution for which $r(t)$ is equal to $1, \gamma t^{\gamma-1}$ and $e^{\gamma t}$ respectively. These are the most widely and frequently used survival models and their resulting Q-functions as defined in (4.1) are not included in the classes considered by either Ford, Torsney and Wu (1992) or Hedayat, Zhong and Nie (2004) (see, for example, section 4.4.3 for the exponential-based model).

Examples of models that do not satisfy our assumptions but are included in the class defined by Yang and Stufken (2009) are the logistic, probit and double exponential models. This is because Yang and Stufken (2009) assumptions on their Q-function are somewhat more general, although not as easy to verify, than ours and therefore their class includes more generalised linear models. For example, the corresponding Q-functions for the logistic, probit and double exponential models are all even functions whereas we concentrate only on strictly increasing functions on $\mathbb{R}$.

However, our assumptions hold but those of Yang and Stufken (2009) do not for certain accelerated failure time models with two failure modes. This corresponds to a situation where the severity of the conditions that the subjects in the experiment are put under changes at a certain point resulting in two different modes of failure. Therefore the type of failure time distribution differs between modes. An example of this is the case where the failure time distribution changes from a Gamma with shape parameter 2 to an exponential distribution depending on the sign of θ. At $\theta = 0$ the resulting Q-function is not three times continuously differentiable as required by Yang and Stufken (2009). This is proven in Appendix A section A.1.5.

4.2 Locally D-optimal designs

In what follows we give the definition of a D-optimal design and the general equivalence theorem for D-optimality for models with information matrix (4.1). We then solve the design problem for binary and continuous design spaces, with Theorem 2 being the main result of this section.

4.2.1 The criterion

If we are interested in estimating both of the model parameters α and β the optimality criterion we should use for the construction of the designs is D-optimality. A D-optimal design maximises the determinant of the information matrix $M(\xi, \alpha, \beta)$ with respect to the design ξ. It therefore minimises the volume of the confidence ellipsoid for the parameter estimators and so it makes the estimators as precise as possible. That is, a design ξ^* is D-optimal if

$$\xi^* = \arg\max_{\xi} |M(\xi, \alpha, \beta)|.$$

A useful tool for characterising D-optimal designs and for checking the D-optimality of a candidate design is the general equivalence theorem (see, for example, Silvey (1980)). The following theorem presents the general equivalence theorem for models in the class we consider.

Theorem 1. *A design ξ^* is locally D-optimal for a model with information matrix (4.1) if the inequality*

$$d(\xi^*, \alpha, \beta) = tr\{M^{-1}(\xi^*, \alpha, \beta) I(x, \alpha, \beta)\} \leq 2,$$

holds for all $x \in \mathcal{X}$, with equality in the support points of ξ^*.

4.2.2 Binary design space

To allow estimation of both parameters a design must have at least two support points. In the case of a binary design space $\mathcal{X} = \{0, 1\}$ this means that both points, 0 and 1, are support points of the locally D-optimal design. From Lemma 5.1.3 in Silvey (1980), it follows that for any model with information matrix of the form (4.1) the D-optimal design with as many support points as there are model parameters, has equal weights. Therefore the locally D-optimal design ξ^* on the design space $\mathcal{X} = \{0, 1\}$ is

$$\xi^* = \begin{Bmatrix} 0 & 1 \\ 0.5 & 0.5 \end{Bmatrix}.$$

4.2.3 Continuous design space

We now consider design spaces that are intervals, that is, $\mathcal{X} = [u, v]$. For a continuous explanatory variable the D-optimality criterion is invariant under linear transformations of the design space (see, for example, Silvey (1980)) and we can therefore without loss of generality consider the design space $\mathcal{X} = [0, 1]$. The locally D-optimal design for given α and β on an arbitrary interval $[u, v]$ can be obtained from the locally D-optimal design on the interval $[0, 1]$ for parameter values $\tilde{\alpha} = \alpha + \beta u$ and $\tilde{\beta} = \beta(v - u)$ by transforming its support points $\tilde{x_i}$ via $x_i = u + (v - u)\tilde{x_i}$.

From Caratheodory's theorem (see, for example, Silvey (1980)), there exists a D-optimal design with at most three support points. Lemma 1 shows that this number can be further reduced. Its proof is given in section A.1.6 of Appendix A and it modifies an idea of Biedermann-Dette-Zhu.

Lemma 1. *Let $\beta \neq 0$ and conditions (a)-(c) be satisfied. Then the locally D-optimal design for a model with information matrix (4.1) is unique and has two equally weighted support points.*

We now present the main result of this section, that is, an analytical characterisation of locally D-optimal designs for models included in the class under consideration.

Theorem 2. *Let conditions (a)-(d) be satisfied.*
(a) If $\beta > 0$, the design

$$\xi^* = \begin{Bmatrix} x_0^* & 1 \\ 0.5 & 0.5 \end{Bmatrix}$$

is locally D-optimal on $\mathcal{X} = [0, 1]$, where $x_0^* = 0$ if $\beta < 2Q(\alpha)/Q'(\alpha)$. Otherwise, x_0^* is the unique solution of the equation $\beta(x_0 - 1) + 2Q(\alpha + \beta x_0)/Q'(\alpha + \beta x_0) = 0$.
(b) If $\beta < 0$, the design

$$\xi^* = \begin{Bmatrix} 0 & x_1^* \\ 0.5 & 0.5 \end{Bmatrix}$$

is locally D-optimal on $\mathcal{X} = [0, 1]$, where $x_1^* = 1$ if $\beta > -2Q(\alpha + \beta)/Q'(\alpha + \beta)$. Otherwise, x_1^* is the unique solution of the equation $\beta x_1 + 2Q(\alpha + \beta x_1)/Q'(\alpha + \beta x_1) = 0$.

Proof. Here we only give a sketch of the proof for part (a). The proof of part (b) follows along the same lines using symmetry arguments and is presented in detail in section A.1.7. of Appendix A.

Let $\beta > 0$. Since conditions (a)-(c) are satisfied Lemma 1 can be used. For the locally D-optimal design equally supported at points $x_0^*, x_1^* \in [0, 1]$, where $x_0^* < x_1^*$, the determinant of (4.1) is increasing with x_1^*, regardless of the value of x_0^*. Therefore it is maximised for $x_1^* = 1$ and it remains to maximise the function

$$g_2(\alpha + \beta x_0) = Q(\alpha + \beta x_0)(x_0 - 1)^2.$$

Using condition (d), $g_2(\alpha + \beta x_0)$ has exactly two turning points on $(-\infty, 1]$, one of which is a minimum at $x_0 = 1$, hence the other one must be a maximum. If this maximum is attained outside the design space, $g_2(\alpha + \beta x_0)$ is maximised at $x_0 = 0$, which will then be the second support point x_0^* of the locally D-optimal design. This occurs if and only if $(\partial/\partial x_0)g_2(\alpha+\beta x_0) < 0$ at $x_0 = 0$, which is equivalent to $\beta < 2Q(\alpha)/Q'(\alpha)$. Otherwise, the point at which the maximum is attained will be the smaller support point x_0^*. This is found by solving $(\partial/\partial x_0)g_2(\alpha + \beta x_0) = 0$, which is equivalent to solving $\beta(x_0 - 1) + 2Q(\alpha + \beta x_0)/Q'(\alpha + \beta x_0) = 0$. □

Theorem 2 provides a complete classification of locally D-optimal designs. Depending on the sign of β, one of the support points is always fixed at one of the boundaries of the design space and according to some easily verifiable conditions on the parameters the design problem has either been reduced to an optimisation problem in one variable or been solved completely.

Russell et al. (2009) consider the construction of locally D-optimal designs for Poisson regression with log-linear link which is included in our class of models for $Q(\theta) = e^\theta$ as shown in section 4.1.2. Using Theorem 2 we have that for $|\beta| \geq 2$ the equally weighted D-optimal support points are $\{1 - 2/\beta, 1\}$ if $\beta > 0$ and $\{0, -2/\beta\}$ if $\beta < 0$, and therefore the parameter α does not affect the optimal choice of design. This matches the results of the main theorem in Russell et al. (2009) for a design space $\mathcal{X} = [0, 1]$ and one explanatory variable. The corresponding locally D-optimal design on the transformed design space $\Theta = \alpha + \beta\mathcal{X}$ is equally supported at $\{\alpha + \beta - 2, \alpha + \beta\}$ if $\beta > 0$ and $\{\alpha - 2, \alpha\}$ if $\beta < 0$ for $|\beta| \geq 2$, whereas if $|\beta| < 2$ then the support points are $\{\alpha, \alpha + \beta\}$ and $\{\alpha + \beta, \alpha\}$ for positive and negative β-values respectively. Therefore, our designs are also in accordance with the results of Ford, Torsney and Wu (1992) for models with information matrix (4.1) and Q-function of the form e^θ (see Table 3 in Ford, Torsney and Wu (1992)).

4.3 Locally c-optimal designs for estimating β

As in the previous section we first present the optimality criterion and the corresponding general equivalence theorem and we also give the motivation for the use of c-optimality for estimating the parameter β. Some general results are then discussed and the cases of a binary and a continuous design space are again investigated separately. A complete classification of locally c-optimal designs for estimating β is given in Theorem 4.

4.3.1 The criterion

Often interest centres in estimating the parameter β while treating α as a nuisance parameter. The motivation for this choice of parameter comes from the parametrisation of proportional hazards models given in (4.2). Under this parametrisation the parameter α relates to the baseline hazard whereas β describes the effect of the explanatory variable x and is therefore reasonable for β to be the main parameter of interest.

In this case the appropriate criterion to use is c-optimality for β which minimises the asymptotic variance of the maximum likelihood estimator $\hat{\beta}$. Thus a design ξ^* is c-optimal for β if the vector $(0 \;\; 1)^T$ is in the range of $M(\xi^*, \alpha, \beta)$ and

$$\xi^* = \arg\min_{\xi} (0 \;\; 1)\, M^{-}(\xi, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{4.3}$$

where $M^{-}(\xi, \alpha, \beta)$ is a generalised inverse of the matrix $M(\xi, \alpha, \beta)$. The corresponding general equivalence theorem for c-optimality for β and models in the class considered is given below.

Theorem 3. *A design ξ^* is locally c-optimal for estimating β for a model with information matrix (4.1) if the inequality*

$$\left(\sqrt{Q(\alpha + \beta x)} \quad x\sqrt{Q(\alpha + \beta x)}\right) M^{-}(\xi, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \leq (0 \;\; 1)\, M^{-}(\xi, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

holds for all $x \in \mathcal{X}$, with equality in the support points of ξ^*.

4.3.2 General results

We now present some results which are applicable to both a binary and a continuous design space. From Caratheodory's Theorem (see, for example, Silvey (1980)) applied to the Elfving set (see Elfving (1952)), there exists a c-optimal design for β with at most two support points. The following lemma shows that a locally c-optimal design for β for models with information matrix of the form (4.1) is supported on exactly two points.

Lemma 2. *For any choice of α, β ($\beta \neq 0$) and any model with information matrix (4.1) there exists a locally c-optimal design for estimating β with exactly two support points.*

Proof. We assume that there exists a locally c-optimal design for β with only one support point $\tilde{x}$. For estimability we require that $(0 \;\; 1)^T$ is in the range of $M(\xi, \alpha, \beta)$, that is, there exists a vector $\eta = (\eta_1, \eta_2)^T \in \mathbb{R}^2$ such that

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = Q(\alpha + \beta\tilde{x}) \begin{pmatrix} 1 & \tilde{x} \\ \tilde{x} & \tilde{x}^2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \iff \begin{pmatrix} 0 = Q(\alpha + \beta\tilde{x})(\eta_1 + \eta_2\tilde{x}) \\ 1 = Q(\alpha + \beta\tilde{x})\tilde{x}(\eta_1 + \eta_2\tilde{x}) \end{pmatrix}$$

From the first equation we obtain that $Q(\alpha + \beta\tilde{x})\eta_1 = -Q(\alpha + \beta\tilde{x})\eta_2\tilde{x}$. Substituting this into the second equation yields $1 = 0$. Therefore no locally c-optimal design for β with only one support point exists.

□

From Pukelsheim and Torsney (1991) we obtain an expression for the c-optimal weights. That is, for models with information matrix (4.1) a c-optimal design ξ^* for β with support points x_0^* and x_1^*, where $x_0^* < x_1^*$, is given by

$$\xi^* = \left\{ \begin{array}{cc} x_0^* & x_1^* \\ \frac{\sqrt{Q(\alpha+\beta x_1^*)}}{\sqrt{Q(\alpha+\beta x_0^*)}+\sqrt{Q(\alpha+\beta x_1^*)}} & \frac{\sqrt{Q(\alpha+\beta x_0^*)}}{\sqrt{Q(\alpha+\beta x_0^*)}+\sqrt{Q(\alpha+\beta x_1^*)}} \end{array} \right\}. \quad (4.4)$$

Using condition (b) we have that for positive values of the parameter β the function $Q(\alpha+\beta x)$ is increasing with x, whereas it is decreasing for negative β-values. Therefore, from (4.4) we can observe that the c-optimal weight corresponding to the smaller support point x_0^* is greater than the weight corresponding to x_1^* for $\beta > 0$ and smaller for negative values of β.

4.3.3 Binary design space

From Lemma 2 we know that the c-optimal design for β is supported at exactly two points which will be 0 and 1 in the case of a binary design space. The design problem is thus solved completely by also using the expressions given in (4.4) and the c-optimal design ξ^* for estimating β on the design space $\mathcal{X} = \{0, 1\}$ is

$$\xi^* = \left\{ \begin{array}{cc} 0 & 1 \\ \frac{\sqrt{Q(\alpha+\beta)}}{\sqrt{Q(\alpha)}+\sqrt{Q(\alpha+\beta)}} & \frac{\sqrt{Q(\alpha)}}{\sqrt{Q(\alpha)}+\sqrt{Q(\alpha+\beta)}} \end{array} \right\}.$$

It is interesting to note that the popular equal allocation rule which is almost always used in practice when comparing, for example, two methods or treatments leads to a suboptimal design.

4.3.4 Continuous design space

Unlike D-optimality, the c-optimality criterion does not satisfy an invariance property and therefore we find locally c-optimal designs for estimating β on an arbitrary design space $\mathcal{X} = [u, v]$. An analytical characterisation of the locally c-optimal designs for β for models with information matrix of the form (4.1) is provided in Theorem 4. A sketch proof of part (a) is given below and part (b) is proven in Appendix A, section A.1.8.

Theorem 4. *Let conditions (a), (b) and (d1) be satisfied.*
(a) If $\beta > 0$, the design ξ^* with support points x_0^* and v and the optimal weights given in (4.4) is locally c-optimal for β on $\mathcal{X} = [u, v]$, where $x_0^* = u$ if

$$\beta(u-v)+2\frac{Q(\alpha+\beta u)}{Q'(\alpha+\beta u)}\left(1+\frac{\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta v)}}\right)>0. \tag{4.5}$$

Otherwise, x_0^* is the unique solution of the equation

$$\beta(x_0-v)+2\frac{Q(\alpha+\beta x_0)}{Q'(\alpha+\beta x_0)}\left(1+\frac{\sqrt{Q(\alpha+\beta x_0)}}{\sqrt{Q(\alpha+\beta v)}}\right)=0. \tag{4.6}$$

(b) If $\beta<0$, the design ξ^* with support points u and x_1^* and the optimal weights given in (4.4) is locally c-optimal for β on $\mathcal{X}=[u,v]$, where $x_1^*=v$ if

$$\beta(u-v)-2\frac{Q(\alpha+\beta v)}{Q'(\alpha+\beta v)}\left(1+\frac{\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}}\right)<0.$$

Otherwise, x_1^* is the unique solution of the equation

$$\beta(u-x_1)-2\frac{Q(\alpha+\beta x_1)}{Q'(\alpha+\beta x_1)}\left(1+\frac{\sqrt{Q(\alpha+\beta x_1)}}{\sqrt{Q(\alpha+\beta u)}}\right)=0.$$

Proof. Let $\beta>0$ and also let ξ^* be a locally c-optimal design for β which, following Lemma 2, has exactly two support points $x_0^*, x_1^* \in [u,v]$, where $x_0^*<x_1^*$. Substituting the expressions for the c-optimal weights from (4.4), we obtain the objective function to be minimised defined in (4.3), to be given by

$$\tilde{d}(x_0^*,x_1^*):=\left(\frac{1}{\sqrt{Q(\alpha+\beta x_0^*)}}+\frac{1}{\sqrt{Q(\alpha+\beta x_1^*)}}\right)^2\frac{1}{(x_0^*-x_1^*)^2}.$$

Holding x_0^* fixed, $\tilde{d}(x_0^*,x_1^*)$ is decreasing with x_1^* and therefore attains its minimum in $[u,v]$ at the upper bound v of the design space. Now using conditions (a), (b) and $(d1)$ it can be shown that $\tilde{d}(x_0,v)$ has exactly one turning point on $(-\infty,v)$ and so there is at most one turning point in $[u,v]$, which is a minimum since

$$\lim_{x_0\to-\infty}\tilde{d}(x_0,v)=\lim_{x_0\to v}\tilde{d}(x_0,v)=\infty.$$

If this minimum is attained outside $[u,v)$ the lower bound u of the design space is the smaller support point x_0^* of the locally c-optimal design for β. This occurs if and only if $(\partial/\partial x_0)\tilde{d}(x_0,v)>0$ at $x_0=u$, which is equivalent to condition (4.5). Otherwise x_0^* is the unique point where $\tilde{d}(x_0,v)$ is minimised and can be found by solving $(\partial/\partial x_0)\tilde{d}(x_0,v)=0$, which is equivalent to solving (4.6).

□

Using Theorem 4, the locally c-optimal designs for β can be found by minimising just a one-variable function thus reducing the numerical effort substantially. We also note that the optimal weights are always unequal. This contradicts the standard designs used in practice which, as it is shown here, are suboptimal.

For the Poisson regression model with rate $e^{\alpha+\beta x}$, $Q(\theta) = e^{\theta}$ and if we apply the results of Theorem 4 we have that for positive values of the parameter β the locally c-optimal design is supported at $\{u, v\}$ if $\beta(u-v) + 2(1 + e^{\beta(u-v)/2}) > 0$ and at $\{v - 2.56/\beta, v\}$ otherwise. This matches the results in Ford, Torsney and Wu (1992) for the transformed design space $\Theta = \alpha + \beta\mathcal{X}$, which for $\beta > 0$ is equal to $[\alpha + \beta u, \alpha + \beta v]$, stating that the support points of the locally c-optimal designs are $\{\max(\alpha + \beta u, \alpha + \beta v - 2.56), \alpha + \beta v\}$ (see Table 2 in Ford, Torsney and Wu (1992)).

4.4 Application to the exponential regression model

Here we apply the previous results to the exponential regression model in its proportional hazards parametrisation. We first introduce the model and also discuss the case of no censoring. Two censoring mechanisms are considered and it is verified that the model is included in the class we have defined under both scenarios. We then construct locally D- and c-optimal designs for various vectors of parameter values using the analytical characterisations given in Theorems 2 and 4 respectively and identify how the optimal choice of design changes in the presence of censoring. Finally, based on these conclusions we give recommendations to practitioners on which design to use in time to event experiments.

4.4.1 The model

Let $T_1, \ldots, T_n$ be independent random variables indicating the times to event of the n subjects in an experiment of total duration c, with $t_1, \ldots, t_n$ the corresponding observed values. The exponential regression model in its proportional hazards parametrisation is specified by the probability density function

$$f(t_j, x_j) = e^{\alpha+\beta x_j} e^{-t_j e^{\alpha+\beta x_j}}, \quad t_j > 0 \tag{4.7}$$

where $x_j \in \mathcal{X}$, $j = 1, \ldots, n$ is the value of the explanatory variable for the jth subject.

This application is motivated by the fact that model (4.7) is the simplest and one of the most frequently used survival models in practice. An exponential distribution along with the proportional hazards assumption is often assumed for the times until the occurrence of the event of interest. Also the proportional hazards parametrisation avoids the need to constrain the model parameters.

4.4.2 No censoring

The special case of no censoring corresponds to a censoring time $c = \infty$. That is, an experiment running for as long as necessary in order for all times to event to be recorded. From (4.7), the log-likelihood at x_j is

$$l(x_j, \alpha, \beta) = \alpha + \beta x_j - t_j e^{\alpha + \beta x_j},$$

and thus the Fisher information matrix at the point x_j is given by

$$I(x_j, \alpha, \beta) = \begin{pmatrix} E\left(-\frac{\partial^2 l}{\partial \alpha^2}\right) & E\left(-\frac{\partial^2 l}{\partial \alpha \partial \beta}\right) \\ E\left(-\frac{\partial^2 l}{\partial \alpha \partial \beta}\right) & E\left(-\frac{\partial^2 l}{\partial \beta^2}\right) \end{pmatrix} = \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix},$$

using the fact that the times to event follow an exponential distribution with mean $E(T_j) = 1/e^{\alpha+\beta x_j}$. In this case the Fisher information is in fact the same as for the homoscedastic linear model $T_j = \alpha + \beta x_j + \epsilon_j$ for independent identically distributed errors $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$.

It is well known (see, for example, Atkinson, Donev and Tobias (2007)) that the D-optimal design for the homoscedastic linear model is equally supported at the end-points of the design space $\mathcal{X}$. For the c-optimality case we observe that the Fisher information matrix yields (4.1) with $Q(\theta) = 1$. The objective function defined in (4.3) for model (4.7) is then $1/(x_1 - x_2)^2$, that is the inverse of the determinant of the information matrix. Therefore, the locally D-optimal design allocating equal weight to the end points of the design space $\mathcal{X}$ is also locally c-optimal for β in this case.

4.4.3 Right-censoring

We now assume that some of the observations are right-censored. That is, a subject's actual event time cannot be observed if it exceeds the subject's censoring time.

The first mechanism we consider that can result in right-censored observations is Type-I censoring under which the censoring time is common for all the subjects. This occurs, for example, if all the subjects are recruited at the same time in an experiment of predetermined total duration which will therefore be the fixed and common censoring time. On the other hand, in the case of random censoring the censoring time is possibly different for each subject and independent of the corresponding time to event. We consider the following type of random censoring. Suppose the duration of the experiment is fixed, but subjects are recruited randomly within that time interval. Therefore, the time of entrance for each subject is uniformly distributed and if the desired event has not been observed for a subject by the end of the experiment its corresponding observation is right-censored.

In the presence of Type-I censoring what we actually observe for each subject under investigation is $Y_j = min\{T_j, c\}$. Therefore, if the event of interest has not occurred by the end of the experiment, that is, by time c, the observation is right-censored. Let T_j follow model (4.7). Then

$$E(Y_j) = \int_0^c y e^{\alpha+\beta x_j} e^{-ye^{\alpha+\beta x_j}}\, dy + cP(Y_j = c) = (1 - e^{-ce^{\alpha+\beta x_j}})/e^{\alpha+\beta x_j}, \qquad (4.8)$$

and the log-likelihood at x_j is $l(\alpha, \beta, x_j) = \delta_j(\alpha + \beta x_j) - y_j e^{\alpha+\beta x_j}$, where δ_j is an event indicator which is zero if y_j is a right-censored observation and unity otherwise. Hence the Fisher information at x_j is

$$I(x_j, \alpha, \beta) = (1 - e^{-ce^{\alpha+\beta x_j}}) \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix}.$$

This yields (4.1) with $Q(\theta) = (1 - e^{-ce^{\theta}})$ which satisfies conditions (a)-(d) and (d1) since model (4.7) is a special case of parametric proportional hazards models of the form (4.2) discussed in section 4.1.2.

For random censoring we assume that the subjects enter the experiment at random times $Z_j \in [0, c]$, $j = 1, \ldots, n$, where Z_j is independent of the time to event T_j. Hence the censoring times $C_j = c - Z_j$, $j = 1, \ldots, n$ are also random. We further assume that $Z_1, \ldots, Z_n$ follow a uniform distribution on $[0, c]$, thus $C_1, \ldots, C_n$ also have a uniform distribution on $[0, c]$ with probability density function $f_c(c_j) = 1/c$. Here we observe $Y_j = min\{T_j, C_j\}$ where $E(Y_j|C_j = c_j)$ is given by the right hand side of (4.8) with c replaced by c_j. Thus

$$\begin{aligned} E(Y_j) = E(E(Y_j|C_j = c_j)) &= \int_0^c \frac{(1 - e^{-c_j e^{\alpha+\beta x_j}})}{ce^{\alpha+\beta x_j}}\, dc_j \\ &= \left(ce^{\alpha+\beta x_j} + e^{-ce^{\alpha+\beta x_j}} - 1\right) / ce^{2(\alpha+\beta x_j)}. \end{aligned}$$

The log-likelihood at x_j is $l(x_j, \alpha, \beta) = \delta_j(-\log c + \alpha + \beta x_j) - y_j e^{\alpha+\beta x_j}$, where δ_j is zero if y_j is a right-censored observation and unity otherwise. Hence the Fisher information at point x_j is given by

$$I(x_j, \alpha, \beta) = \frac{\left(ce^{\alpha+\beta x_j} + e^{-ce^{\alpha+\beta x_j}} - 1\right)}{ce^{\alpha+\beta x_j}} \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix}.$$

Again this is of the form (4.1) for $Q(\theta) = 1 + \frac{\left(e^{-ce^{\theta}} - 1\right)}{ce^{\theta}}$. Assumptions (a)-(d) and (d1) hold as a special case of model (4.2) subject to random censoring (see section 4.1.2).

4.4.4 Locally optimal designs

As mentioned in section 4.2.2, in the case of a binary design space $\mathcal{X} = \{0, 1\}$ the locally D-optimal design is always equally supported at points 0 and 1 regardless of the parameter values. We therefore consider the continuous design space $\mathcal{X} = [0, 1]$ and use Theorem 2, presented in section 4.2.3, for the construction of the designs. Following this, the two D-optimal weights are always equal and the corresponding support points are found by solving an optimisation problem in just one variable. Tables 4.1 and 4.2 give the support points of locally D-optimal designs for model (4.7) in the cases of Type-I and random censoring respectively. Both positive and negative β-values are considered and small values of ce^{α} correspond to large percentages of censoring.

Table 4.1: Support points for some selected locally D-optimal designs for model (4.7) under Type-I censoring

ce^{α}	β					
	-2.3	-2.2	-2.1	2.1	2.2	2.3
0.1	(0,0.88)	(0,0.92)	(0,0.96)	(0,1)	(0.04,1)	(0.08,1)
0.01	(0,0.87)	(0,0.91)	(0,0.95)	(0.04,1)	(0.09,1)	(0.12,1)
0.001	(0,0.87)	(0,0.91)	(0,0.95)	(0.05,1)	(0.09,1)	(0.13,1)

Table 4.2: Support points for some selected locally D-optimal designs for model (4.7) under random censoring

ce^{α}	β					
	-2.3	-2.2	-2.1	2.1	2.2	2.3
0.1	(0,0.87)	(0,0.91)	(0,0.96)	(0.01,1)	(0.06,1)	(0.09,1)
0.01	(0,0.87)	(0,0.91)	(0,0.95)	(0.04,1)	(0.09,1)	(0.13,1)
0.001	(0,0.87)	(0,0.91)	(0,0.95)	(0.05,1)	(0.09,1)	(0.13,1)

The above results indicate that censoring affects the optimal choice of design for model (4.7) in both censoring scenarios, which produce similar results. When the parameter β is positive the probability of occurrence of the event of interest increases with the explanatory variable x. Hence the point $x = 1$ is more informative and is always included in the locally D-optimal design. We also observe that for positive β-values the smaller support point of the design is greater than zero. This is because the possibility of censoring and therefore the variance at $x = 0$ is greater. The bigger the β-values and/or the smaller the ce^{α}-values are, the bigger the variance at $x = 0$ is and so the smaller support point of the locally D-optimal design is chosen to be further away from zero. In the opposite case of $\beta < 0$ the locally D-optimal design is always supported at $x = 0$ and tends to include a point smaller than one as the larger support point.

We note that the β-values used in Tables 4.1 and 4.2 correspond to large effects of the explanatory variable. In a medical application x could be a measure of the dose of a drug with time to relief of symptoms, as the response time. There would be a large effect size if the time to symptom relief is sensitive to dose. In an engineering example x could be a measure of stress on a system and the time to failure of the system, the time of interest. Then the effect size would be large, if the failure time was highly stress dependent.

The construction of locally c-optimal designs for estimating β for model (4.7) is facilitated by our results given in Theorem 4, see section 4.3.4, assuming a continuous design space $\mathcal{X} = [0, 1]$. For the parameter values chosen here the two support points of the designs are always 0 and 1 and therefore the locally c-optimal designs for β on the binary design space, $\mathcal{X} = \{0, 1\}$, and the continuous designs space, $\mathcal{X} = [0, 1]$ coincide. The c-optimal weights on $x = 0$ are then found using (4.4), given in section 4.3.2, for Type-I and random censoring separately and are presented in Tables 4.3 and 4.4.

Table 4.3: Weights on $x = 0$ for some selected locally c-optimal designs for β for model (4.7) under Type-I censoring

ce^{α}	β							
	-1	-0.7	-0.4	-0.1	0.1	0.4	0.7	1
0.1	0.39	0.42	0.45	0.49	0.51	0.55	0.58	0.61
0.01	0.38	0.41	0.45	0.49	0.51	0.55	0.59	0.62
0.001	0.38	0.41	0.45	0.49	0.51	0.55	0.59	0.62

Table 4.4: Weights on $x = 0$ for some selected locally c-optimal designs for β for model (4.7) under random censoring

ce^{α}	β							
	-1	-0.7	-0.4	-0.1	0.1	0.4	0.7	1
0.1	0.38	0.42	0.45	0.49	0.51	0.55	0.58	0.62
0.01	0.38	0.41	0.45	0.49	0.51	0.55	0.59	0.62
0.001	0.38	0.41	0.45	0.49	0.51	0.55	0.59	0.62

We observe that even for small values for the parameter β the c-optimal weights are not equal, unlike locally D-optimal designs which are always equally supported. In particular, both for Type-I and random censoring the locally c-optimal design for β allocates more subjects to point $x = 0$ for $\beta > 0$ and less in the case of negative values for the parameter β. That is, the design puts more weight at the experimental point where censoring is more likely so that the variance is minimised. Therefore, even for small β-values and thus small effects of the explanatory variable, the standard design allocating half the subjects at point $x = 0$ and the rest at $x = 1$ is not optimal for either of the censoring scenarios.

We also note that there is an obvious symmetry in the c-optimal weight values for equal departures of the β-values from the trivial case of $\beta = 0$. In particular, for a positive β-value the c-optimal weight at $x = 0$ is equal to that at point $x = 1$ for the corresponding negative value of β equally away from $\beta = 0$.

4.4.5 Recommendations

Based on the results presented in the previous section we can give advice to practitioners on how to plan an experiment involving survival models. In particular, if a design is required before the commencement of a time to event experiment we would recommend the use of a c-optimal rather than a D-optimal design and the c-optimality criterion should be chosen for the estimation of the model parameter corresponding to the explanatory variable effect. This is because the ultimate goal of such time to event experiments is to explain how a particular covariate, which might be, for example, a method or treatment, is related to the time to the event under investigation. It is therefore reasonable to focus on estimating only the covariate parameter even if the second model parameter is not known, although in many practical situations some information is usually available (see Chapter 5 for discussion).

Under the c-optimality criterion it is evident, both from the analytical characterisations of locally c-optimal designs presented in section 4.3 as well as the numerical results given in Tables 4.3 and 4.4, that the standard design allocating an equal number of subjects at the end-points of the design space is not optimal in the presence of censoring. This can be explained by the fact that the amount of information is lower at the experimental point where the probability of censoring is greater and therefore the optimal design puts more weight at that point in order to balance this out. For our recommendation we also take into account that a practitioner has knowledge about the type of the explanatory variable effect and so the sign of the covariate parameter. For example, a new treatment will go under study only when it is expected to be superior to the one currently in use which corresponds to a negative sign for the treatment parameter (assuming that long times to event are preferable). Therefore a locally c-optimal design allocating more than half of the subjects to the experimental point where censoring is more likely to occur will be at least better than the standard design, if not optimal.

Chapter 5

Parameter-robust designs

Due to the nonlinearity of the models considered the optimal designs found in the previous chapter depend on the model parameters which are unknown in practice. Hence there is the need to overcome this dependence and construct designs which are robust to parameter misspecifications, that is, designs which estimate the parameters of interest with high precision and therefore perform well, even when there is imperfect knowledge of the true parameter values. In many practical situations, some information about the parameters, such as a range of plausible values, can be provided by the experimenter. In particular, throughout this chapter we fix the value of the constant term α, whereas for the parameter β a range of values is specified. We further assume that the experimenter has no preference for specific β-values.

The scenario described above is motivated by the interpretation of the parameters involved in a model under a proportional hazards parametrisation (see equation (4.2) for a general form of the hazard function). Consider, for example, a clinical trial where patients are randomised to receive a standard or a new treatment. The expression e^{α} represents the hazard of the event occurring for patients on the standard treatment and having in mind that a standard treatment which is in use has been previously investigated in depth, a reasonable approximation to the value of α may be available. Moreover, the parameter β describes how the risk of the event occurring changes according to the new treatment, and assuming large times to event are preferred it has a negative value when the new treatment is superior to the standard one. Therefore, the experimenter can specify a range of β-values for a clinically relevant important improvement with the new treatment.

In what follows, we present the two design strategies for the construction of parameter robust designs for models in the class introduced in section 4.1.1 and under the parameter information scenario discussed. In particular, we first investigate standardised maximin D- and c-optimal designs as well as cluster designs, the construction of which is facilitated by our analytical results on locally optimal designs. Then the robustness of these designs and of locally optimal designs is assessed by comparing their efficiencies when the parameters have been misspecified. We use the exponential regression model in its proportional hazards parametrisation defined in section 4.4.1 throughout the robustness analysis and we illustrate that, unlike traditional designs currently in use, our designs perform well across a broad range of scenarios.

5.1 Standardised maximin optimal designs

Here we consider the construction of designs using a maximin efficiency approach where a design is evaluated according to its performance in the worst possible case. Unlike the Bayesian optimal designs approach, under this concept there is no need for a prior distribution to be specified for the model parameters, thereby avoiding a step that can be difficult in practice.

The calculation of our designs is based on standardised optimality criteria introduced by Dette (1997). The main advantage of these criteria is that they satisfy an invariance property for linear transformations of the design space similar to the one for the D-optimality criterion. Therefore, the design on the transformed space can be obtained by scaling the support points according to the transformation of the design space while leaving the weights unchanged.

Following Dette (1997), we seek designs that maximise the minimum efficiency with respect to the locally optimal designs found over a certain range of parameter values. This allows us to construct designs which protect against the worst case scenario for the parameter misspecification. As pointed out in Dette and Biedermann (2003) the standardised maximin optimal designs are usually found by optimisation in the subclass of minimally supported designs. Using our results proved in the previous chapter, the locally optimal designs for models with information matrix of the form (4.1) and satisfying conditions (a)-(d) and (d1), always have exactly two support points and therefore we search for the best performing two-point design.

According to the parameter information scenario previously discussed, we assume that the true value of the parameter α is known whereas the parameter β takes values in the interval $[\beta_0, \beta_1]$ with $\beta_0 < \beta_1$. A standardised maximin optimal two-point design maximises the criterion

$$\Phi(\xi) = \min \left\{ eff(\xi) \;\middle|\; \beta \in [\beta_0, \beta_1] \right\},$$

in the class of all two-point designs, where the efficiency $eff(\xi)$ differs according to which combination of model parameters we are interested in estimating and therefore which optimality criterion we consider.

We now define the two standardised maximin criteria for D-and c-optimality and provide analytical characterisations of the designs found by maximising these criteria in Theorems 5 and 6 respectively.

5.1.1 Standardised maximin D-optimal designs

The D-efficiency of an arbitrary design ξ is defined as

$$eff_D(\xi) = \left(\frac{|M(\xi, \alpha, \beta)|}{|M(\xi^*_\beta, \alpha, \beta)|} \right)^{1/2}, \tag{5.1}$$

where ξ^*_β is the locally D-optimal design on the design space $\mathcal{X}$ (see, for example, Atkinson, Donev and Tobias (2007)). The square root of the ratio of determinants is taken so that the efficiency has the dimensions of a ratio of variances. Therefore, if a design with 50% efficiency is used it will require double the replicates for it to be able to estimate the model parameters with similar variances as for the optimal design. Following Dette (1997), a design ξ^* maximising the criterion

$$\Phi(\xi) = \min \left\{ \left(\frac{|M(\xi,\alpha,\beta)|}{|M(\xi^*_\beta,\alpha,\beta)|} \right)^{1/2} \,\middle|\, \beta \in [\beta_0, \beta_1] \right\},$$

in the class of all two-point designs is called a standardised maximin D-optimal two-point design. Thus, this design maximises the worst D-efficiency over the β-values in the interval $[\beta_0, \beta_1]$.

We briefly discuss the case of a binary design space $\mathcal{X} = \{0, 1\}$ for which, as is shown in section 4.2.2, the locally D-optimal design is equally supported at points 0 and 1 regardless of the parameter values. So no further investigation needs to be done and the standardised maximin D-optimal two-point design also has 0 and 1 as its support points with equal weights.

For an interval design space $\mathcal{X} = [0, 1]$, Lemma 1 in section 4.2.3 states that for a given set of parameter values the locally D-optimal design is unique and always equally supported at exactly two points which are then classified according to some conditions on the parameters in Theorem 2. The following theorem is our main result of this section and gives the standardised maximin D-optimal two-point design for models with information matrix of the form (4.1) when a range of negative β-values is provided.

Theorem 5. *Let $\beta \in [\beta_0, \beta_1]$ where $\beta_1 < 0$, α be fixed and assumptions (a)-(d) and (d1) be satisfied. The standardised maximin D-optimal two-point design on $[0, 1]$ is equally supported at points 0 and x^*_1 where $x^*_1 = 1$ if $\beta_0 > -2Q(\alpha + \beta_0)/Q'(\alpha + \beta_0)$. Otherwise x^*_1 is the solution of the equation*

$$Q(\alpha + \beta_0 x)Q(\alpha + \beta_1 x_{\beta_1})x^2_{\beta_1} = Q(\alpha + \beta_1 x)Q(\alpha + \beta_0 x_{\beta_0})x^2_{\beta_0}, \tag{5.2}$$

where x_{β_0}, x_{β_1} are the solutions of the equation $\beta x + 2\frac{Q(\alpha+\beta x)}{Q'(\alpha+\beta x)} = 0$ for β_0 and β_1 respectively.

Proof. Using part (b) of Theorem 2 presented in section 4.2.3, for any β-value in the interval $[\beta_0, \beta_1]$ the corresponding locally D-optimal design ξ^*_β is equally supported at 0 and x_β where $x_\beta = 1$ if $\beta + 2Q(\alpha+\beta)/Q'(\alpha+\beta) > 0$. Otherwise, x_β satisfies the equation

$$\beta x_\beta + 2\frac{Q(\alpha+\beta x_\beta)}{Q'(\alpha+\beta x_\beta)} = 0.$$

From Silvey (1980) the D-optimal weights of a two-point design under a two-parameter model must be equal. Therefore, for models with information matrix of the form (4.1), the D-efficiency, defined in (5.1), of a two-point design $\xi_{\{x_0,x_1\}}$ equally supported at points $x_0, x_1 \in [0,1]$ with $x_0 < x_1$ is given by

$$eff_D\left(\xi_{\{x_0,x_1\}}\right) = \left\{\frac{Q(\alpha+\beta x_0)Q(\alpha+\beta x_1)(x_0-x_1)^2}{Q(\alpha)Q(\alpha+\beta x_\beta)x_\beta^2}\right\}^{1/2}.$$

Using conditions (a) and (b) it is easy to show that $eff_D\left(\xi_{\{x_0,x_1\}}\right) < eff_D\left(\xi_{\{0,x_1\}}\right)$ for all $x_0 \in [0, x_1)$. Hence the best two-point design is supported at 0 and the standardised maximin D-optimality criterion reduces to

$$\Phi(\xi) = \min\left\{(u(x,\beta))^{1/2} := \left(\tfrac{Q(\alpha+\beta x)x^2}{Q(\alpha+\beta x_\beta)x_\beta^2}\right)^{1/2} \;\middle|\; \beta \in [\beta_0,\beta_1]\right\}.$$

Now using condition (d1) the function $w(\beta) := \beta + 2Q(\alpha+\beta)/Q'(\alpha+\beta)$ is increasing with β. If $w(\beta_0) > 0$, that is, if $\beta_0 > -2Q(\alpha+\beta_0)/Q'(\alpha+\beta_0)$, then $w(\beta) > 0$ for all $\beta \in [\beta_0, \beta_1]$ and so the locally D-optimal design for any $\beta \in [\beta_0,\beta_1]$ is equally supported at points 0 and 1. Therefore, the standardised maximin D-optimal two-point design on $[0,1]$ is also supported at 0 and 1 with equal weights and this completes the proof for the first part of Theorem 5.

In the case of $w(\beta_0) \leq 0$ the following statement holds and is proven in section B.1.1 of Appendix B

(i) For fixed $0 < x \leq 1$, the function $\beta \to u(x,\beta)$ is unimodal.

Hence $u(x,\beta)$ is minimised at β_0 or β_1 and the standardised maximin design can be found by maximising

$$\Phi(\xi) = \min\left\{u(x,\beta_0), u(x,\beta_1)\right\}.$$

This maximisation can be divided into maximisation over the sets

$$M_< := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) < u(x,\beta_1)\right\}$$
$$M_> := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) > u(x,\beta_1)\right\}$$
$$M_= := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) = u(x,\beta_1)\right\}.$$

In section B.1.2 of Appendix B we show that

(ii) The standardised maximin D-optimal two-point design $\xi^*_{\{0,x\}}$ is in the set $M_=$,

and therefore it can be found by solving $u(x, \beta_0) = u(x, \beta_1)$ which is equivalent to solving

$$Q(\alpha + \beta_0 x)Q(\alpha + \beta_1 x_{\beta_1})x^2_{\beta_1} = Q(\alpha + \beta_1 x)Q(\alpha + \beta_0 x_{\beta_0})x^2_{\beta_0}.$$

□

Explicit characterisations of standardised maximin D-optimal designs are also provided, for example, by Dette and Biedermann (2003) for the Michaelis-Menten model. However, to the best of our knowledge, Theorem 5 is the first analytical characterisation of standardised maximin D-optimal designs in a situation where the locally D-optimal designs are not available in closed form.

Based on an easily verifiable condition on the given model parameter values, the standardised maximin D-optimal design is either immediately determined or can be found using an optimisation in just one variable. Therefore, Theorem 5 reduces the numerical effort for design search substantially. We also note that Theorem 5 applies only for negative β-values. The proof used in this case is not applicable when $\beta > 0$ since the solution x_β of the equation

$$\beta(x - 1) + 2\frac{Q(\alpha + \beta x)}{Q'(\alpha + \beta x)} = 0,$$

is concave for positive β-values. Therefore the function $\beta \to u(x, \beta)$ is not unimodal for fixed $0 \le x < 1$ and this is a topic for further investigation.

5.1.2 Standardised maximin c-optimal designs

Following Atkinson, Donev and Tobias (2007) the c-efficiency for estimating the parameter β of an arbitrary design ξ is given by

$$eff_c(\xi) = \frac{(0 \;\; 1)\, M^-(\xi^*_\beta, \alpha, \beta)\binom{0}{1}}{(0 \;\; 1)\, M^-(\xi, \alpha, \beta)\binom{0}{1}}, \tag{5.3}$$

where M^- is a generalised inverse of the information matrix M and ξ^*_β is the locally c-optimal design for estimating β on the design space $\mathcal{X}$. By definition (see, for example, Atkinson, Donev and Tobias (2007)), c-optimal designs for β minimise the asymptotic variance of the estimator $\hat{\beta}$ which is proportional to $(0 \;\; 1)\, M^-(\xi, \alpha, \beta)\binom{0}{1}$. Therefore the above expression of the c-efficiency is already in terms of a ratio of variances and the standardised maximin c-optimal criterion for estimating β is

$$\Phi(\xi) = \min\left\{ \frac{(0 \;\; 1)M^-(\xi^*_\beta,\alpha,\beta)\binom{0}{1}}{(0 \;\; 1)M^-(\xi,\alpha,\beta)\binom{0}{1}} \;\middle|\; \beta \in [\beta_0, \beta_1] \right\}.$$

A design ξ^* with two support points maximising this criterion among all two-point designs and hence maximising the worst c-efficiency over $\beta \in [\beta_0, \beta_1]$, is called a standardised maximin c-optimal two-point design for estimating β.

For the binary design space $\mathcal{X} = \{0, 1\}$, the locally c-optimal design for β is supported at points 0 and 1 and depends on the model parameters through the optimal weights (see section 4.3.3). Theorem 6 provides an analytical characterisation of the standardised maximin c-optimal two-point design for β on $\mathcal{X} = \{0, 1\}$ for models with information matrix of the form (4.1).

Theorem 6. *Let $\beta \in [\beta_0, \beta_1]$, α be fixed and assumptions (a), (b) and (d1) be satisfied. Also let the design space to be binary, that is $\mathcal{X} = \{0, 1\}$. The standardised maximin c-optimal two-point design for β on $\mathcal{X}$ is*

$$\xi^* = \begin{Bmatrix} 0 & 1 \\ \omega^* & 1-\omega^* \end{Bmatrix},$$

where $\omega^* = \frac{\omega_{\beta_0}+\omega_{\beta_1}}{2}$ and ω_{β_0} and ω_{β_1} are the optimal weights at point zero for the locally c-optimal design for β given in (4.4), for β_0 and β_1 respectively.

Proof. It has been shown in section 4.3.3 that in the case of the binary design space $\mathcal{X} = \{0, 1\}$ and for models with information matrix of the form (4.1) the locally c-optimal design for estimating β, ξ^*_β, allocates a proportion ω_β of observations at point 0 and a proportion $1-\omega_\beta$ of observations at 1, where the optimal weights, defined in (4.4), are given by

$$\omega_\beta = \frac{\sqrt{Q(\alpha+\beta)}}{\sqrt{Q(\alpha)}+\sqrt{Q(\alpha+\beta)}},\ 1-\omega_\beta = \frac{\sqrt{Q(\alpha)}}{\sqrt{Q(\alpha)}+\sqrt{Q(\alpha+\beta)}}.$$

Using these expressions for $\omega_\beta, 1-\omega_\beta$, the c-efficiency, defined in (5.3), of a design ξ with support points 0 and 1 and weights ω and $1-\omega$ respectively becomes

$$eff_c(\xi) = \frac{\omega(1-\omega)}{(1-\omega)\omega_\beta^2+\omega(1-\omega_\beta)^2} := u(\omega, \omega_\beta).$$

For fixed ω, taking the first derivative of $u(\omega, \omega_\beta)$ with respect to ω_β and equating it to zero yields $\omega_\beta = \omega$. Furthermore,

$$\left.\frac{\partial^2 u(\omega,\omega_\beta)}{\partial \omega_\beta^2}\right|_{\omega_\beta=\omega} = \frac{-2}{\omega(1-\omega)} < 0.$$

Hence for fixed ω, the function $\omega_\beta \to u(\omega, \omega_\beta)$ is unimodal and so it is minimised at ω_{β_0} or ω_{β_1}. Therefore the standardised maximin c-optimality criterion reduces to maximising

$$\Phi(\xi) = \min\left\{u(\omega, \omega_{\beta_0}), u(\omega, \omega_{\beta_1})\right\}.$$

As before we divide the maximisation of $\Phi(\xi)$ into maximisation over the sets

$$M_< := \left\{\omega \in (0, 1] \;\middle|\; u(\omega, \omega_{\beta_0}) < u(\omega, \omega_{\beta_1})\right\}$$
$$M_> := \left\{\omega \in (0, 1] \;\middle|\; u(\omega, \omega_{\beta_0}) > u(\omega, \omega_{\beta_1})\right\}$$
$$M_= := \left\{\omega \in (0, 1] \;\middle|\; u(\omega, \omega_{\beta_0}) = u(\omega, \omega_{\beta_1})\right\}.$$

If we assume that the maximin c-optimal design is in $M_<$, then we must maximise the function $u(\omega, \omega_{\beta_0})$ and taking its first derivative with respect to ω and equating that to zero yields the solutions $\omega = \omega_{\beta_0}$ and $\omega = \frac{\omega_{\beta_0}}{2\omega_{\beta_0}-1}$. The latter solution is greater than one since $\omega_{\beta_0} < 1$ and so it is rejected. Hence $u(\omega_{\beta_0}, \omega_{\beta_0}) = 1 < u(\omega_{\beta_0}, \omega_{\beta_1})$ which is a contradiction as the efficiency must be greater than or equal to unity. A similar argument for $M_>$ establishes that the standardised maximin c-optimal two-point design for estimating β can be found solving $u(\omega, \omega_{\beta_0}) = u(\omega, \omega_{\beta_1})$ which yields

$$\omega^* = \frac{(\omega_{\beta_0} + \omega_{\beta_1})}{2}.$$

□

Note that no assumptions on the sign of the parameter β have been made and therefore Theorem 6 holds for both positive and negative β-values. As for the locally c-optimal designs, the equal allocation rule which is frequently used in practice leads to a suboptimal design and the optimal weight at $x = 0$ of the standardised maximin c-optimal design is the average of the two locally c-optimal weights at point 0 corresponding to the end-points of the given interval of β-values.

5.2 Cluster designs

The second design strategy we consider for finding parameter robust designs is the construction of cluster designs. These were introduced by Dror and Steinberg (2006) who look at multivariate generalised linear models and find robust designs that take into account uncertainty in the model parameter values, the linear predictor as well as the link function. Their algorithm is based on clustering a set of locally D-optimal designs obtained from a sample of parameter values drawn from the uncertainty space. In particular, the method they propose is to convert this sample of locally D-optimal designs into a set of location vectors and apply a K-means clustering procedure. The centroids of the resulting clusters are then taken to be the support points of the design with equal weights. Dror and Steinberg (2006) verify that this procedure can examine various alternative designs faster, it is simpler and less computationally intensive than other more sophisticated methods such as the Bayesian approach.

Biedermann and Woods (2011) also illustrate that cluster designs perform similarly to and are more easily computed than Bayesian designs through an application to second-harmonic generation experiments. They modify the method proposed by Dror and Steinberg (2006) by allowing the weights of the resulting cluster design to be unequal. Taking into account that the locally optimal weights may differ considerably they apply the clustering to design rather than support points and take the weights to be proportional to the number of points in each cluster.

We find cluster designs for the D- and c-optimality criteria separately, the computation of which is facilitated by our analytical results on locally optimal designs presented in Chapter 4. Under our assumptions about the available information on the model parameters, the construction of cluster designs is as follows. We first draw values uniformly from the range of β-values $[\beta_0, \beta_1]$ provided and calculate the corresponding locally optimal designs. Then the modified clustering algorithm proposed by Biedermann and Woods (2011), provided to us by the authors, is applied since the locally c-optimal designs for models in the class we consider have unequal weights (see section 4.3). The resulting cluster design has the cluster centroids as its support points and the number of support points is allowed to be greater than two when possible. Finally, the weight corresponding to each support point of the design and hence to each cluster, is chosen so that it is proportional to the cluster size.

5.3 Robustness analysis

In this section we assess the robustness of the designs constructed using the two strategies described above and also of the locally optimal designs, by calculating their efficiencies when the parameter values have been misspecified. Again the exponential regression model in its proportional hazards parametrisation defined in section 4.4.1 is used. Type-I censoring and the design space $\mathcal{X} = [0, 1]$ are assumed for demonstration purposes.

For the choice of the uncertainty parameter space we considered the Freireich data (Freireich et al. (1963)). These are data from a two group study to compare a placebo with an active treatment for leukemia. The times to event are the times in months from diagnosis until the death of the patients and are modelled by the exponential-based proportional hazards model. Therefore, we use the maximum likelihood estimates for α and β which are -2.163 and -1.526 respectively.

In what follows we construct the locally D- and c-optimal designs for four sets of parameter values and compare their efficiencies when the parameter values have been misspecified. Then the standardised maximin optimal and cluster designs are computed for a range of β-values and finally all of the above designs are compared separately for the cases of D- and c-optimality. Throughout the efficiency calculations we use a censoring time $c = 30$.

5.3.1 Locally D-optimal designs

We consider locally D-optimal designs ξ_ν for various vectors of parameter values $\boldsymbol{\nu} = (\alpha, \beta)$. The value of the maximum likelihood estimator for α is always used, whereas the β-values are chosen so that they correspond to small ($\boldsymbol{\nu_0}$), medium ($\boldsymbol{\nu_1}$) and large ($\boldsymbol{\nu_3}$) effects of the explanatory variable and also the maximum likelihood estimator value ($\boldsymbol{\nu_2}$). Table 5.1 presents the parameter values used and the corresponding D-efficiencies of the locally D-optimal designs calculated using (5.1) when the vector of parameter values is misspecified .

Table 5.1: D-efficiencies for some selected locally D-optimal designs

	Design			
Parameter vector	ξ_{ν_0}	ξ_{ν_1}	ξ_{ν_2}	ξ_{ν_3}
$\boldsymbol{\nu_0} = (-2.163, -0.1)$	1	1	1	0.900
$\boldsymbol{\nu_1} = (-2.163, -0.405)$	1	1	1	0.905
$\boldsymbol{\nu_2} = (-2.163, -1.526)$	1	1	1	0.946
$\boldsymbol{\nu_3} = (-2.163, -2.623)$	0.992	0.992	0.992	1

The locally D-optimal designs $\xi_{\nu_0}, \xi_{\nu_1}, \xi_{\nu_2}$ corresponding to the first three sets of parameter values are all the 'standard design' supported at 0 and 1, with equal weights, whereas ξ_{ν_3} is equally supported at points 0 and 0.9. We observe that the 'standard design' performs well even in the situation where the true parameter vector is ν_3 in which case its efficiency is equal to 0.992.

Among all the efficiency values the lowest one, 0.900, is obtained if the true vector is $\boldsymbol{\nu_0}$ and the experimenter has misspecified this value as $\boldsymbol{\nu_3}$ and hence used the design ξ_{ν_3}. In other words if the experimenter has used design ξ_{ν_3} assuming a large effect for the explanatory variable when the true effect is actually small, the D-efficiency of design ξ_{ν_3} is 0.9 which is quite satisfying. Hence ξ_{ν_3} seems to be a good alternative to the standard design if, for example, the experimenter does not want to expose the patients at the highest treatment dose.

5.3.2 Locally c-optimal designs

For the same vectors of parameter values used in the previous section, the support points of the locally c-optimal designs are always 0 and 1. The c-optimal weights were found using (4.4), with ω_0 (ω_1) being the weight on the smaller (larger) support point and are shown in Table 5.2. The c-efficiencies of each of the above designs were also calculated using (5.3) when the parameter values are misspecified and are presented in Table 5.3.

Table 5.2: Weights for some selected locally c-optimal designs

	Design			
Weight	ξ_{ν_0}	ξ_{ν_1}	ξ_{ν_2}	ξ_{ν_3}
ω_0	0.498	0.491	0.425	0.323
ω_1	0.502	0.509	0.575	0.677

Table 5.3: c-efficiencies for the locally c-optimal designs of Table 5.2

	Design			
Parameter vector	ξ_{ν_0}	ξ_{ν_1}	ξ_{ν_2}	ξ_{ν_3}
$\boldsymbol{\nu_0} = (-2.163, -0.1)$	1	0.9998	0.9782	0.8772
$\boldsymbol{\nu_1} = (-2.163, -0.405)$	0.9998	1	0.9824	0.8864
$\boldsymbol{\nu_2} = (-2.163, -1.526)$	0.9787	0.9828	1	0.9552
$\boldsymbol{\nu_3} = (-2.163, -2.623)$	0.8908	0.8991	0.9597	1

The design ξ_{ν_2}, which is locally c-optimal for parameter values equal to the maximum likelihood estimator values, has a lowest efficiency of 0.9597 and hence is robust to misspecifications of the parameter space. However, the other three designs do not perform so well in some scenarios. As for the locally D-optimal designs, the lowest efficiency is obtained if the parameter vector is assumed to be ν_3 when in fact its true value is ν_0. This efficiency is equal to 0.8772.

5.3.3 Standardised maximin optimal designs

Using the results presented in section 5.1 we found the standardised maximin D- and c-optimal two-point designs, denoted by ξ_{ν_4} in both cases, for the range $[-2.623, -0.1]$. We note that although here we consider a continuous design space, all the locally c-optimal designs, given in section 5.3.2, are supported at points 0 and 1 for the range of β-values we use and so the result of Theorem 6 can be applied.

The standardised maximin D-optimal design is supported at 0 and 0.993, with equal weights and is locally D-optimal for $\boldsymbol{\nu_4} = (-2.163, -2.380)$, whereas the standardised maximin c-optimal design allocates 41.1% of the observations at the experimental point 0 and the rest at point 1, and is locally c-optimal for $\boldsymbol{\nu_4} = (-2.163, -1.690)$. Compared with the locally D-optimal designs corresponding to the parameter vectors $\boldsymbol{\nu_0}, \ldots, \boldsymbol{\nu_3}$, the minimum (median) efficiency of the standardised maximin D-optimal design is 0.993 (0.993) and 0.969 (0.972) for the standardised maximin c-optimal design. Therefore, both designs perform well across the given parameter space. We note that the minimum efficiency for both of the above designs is obtained when the true parameter vector is $\boldsymbol{\nu_0}$.

5.3.4 Cluster designs

The cluster designs were computed by drawing 1000 β-values uniformly from the interval $[-2.625, -0.1]$. The number of clusters for the locally D-optimal designs was chosen to vary from 2 to 6 and so cluster designs with two up to six support points were constructed. All of the resulting cluster designs give weight 0.5 to the experimental point 0 and very low weight to points not equal to 1. The D-efficiency of each cluster design was also calculated via (5.1) relative to each of the 1000 locally D-optimal designs. The minimum and median efficiencies are found to be the same for all the cluster designs (0.993 and 0.997 respectively) and this may be a result of the very low weight that these designs give to experimental points other than 0 and 1.

The support points of the 1000 locally c-optimal designs are always 0 and 1, hence the cluster design can only have two support points which are the experimental points 0 and 1. Also the clustering here was applied to *design*, rather than *support* points as the support points of the locally c-optimal points have differing weights. The resulting cluster design allocates 43% of the observations at 0 and the rest at 1, and is robust to parameter value misspecifications as the minimum and median efficiencies, found via (5.3) relative to 1000 locally c-optimal designs, are 0.956 and 0.990 respectively.

5.3.5 Comparison of designs

First we compare the performance of the following 10 designs: the locally D-optimal designs $\xi_{\nu_0}, \ldots, \xi_{\nu_3}$, the standardised maximin D-optimal designs ξ_{ν_4} and the cluster designs $\xi_1, \ldots, \xi_5$ with 2 through to 6 support points respectively. The D-efficiency (5.1) of each of the above designs is calculated with respect to each of the 1000 locally optimal designs and the results are summarised in Figure 5.1. Design ξ_{ν_3} was omitted since it was clearly outperformed by the other designs, although it was reasonably efficient (see discussion in section 5.3.1).

From Figure 5.1 we observe that the standardised maximin D-optimal design ξ_{ν_4} is indeed the one with the best minimum efficiency and therefore protects against the worse case scenario. However, it also has a lower median efficiency than the rest of the designs. Another important observation is that the cluster designs $\xi_2, \ldots, \xi_5$ with more than two support points perform similarly to the two-point cluster design ξ_1. Therefore, any one of them can be used instead of the two-point design which also allows us to check for the lack of fit of the model. Finally all five cluster designs and the standardised maximin D-optimal design perform well for the parameter space provided and so they are good alternatives to the locally D-optimal designs.

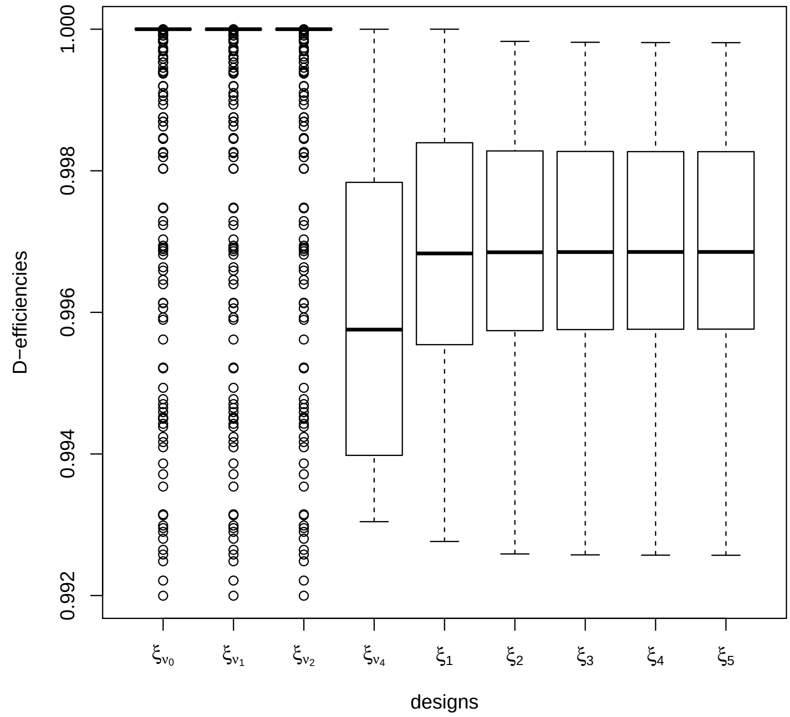


Figure 5.1: Boxplots of D-efficiencies calculated for 9 designs for 1000 parameter vectors

The comparison of the c-optimal designs is shown in Figure 5.2. The designs compared here are the locally c-optimal designs $\xi_{\nu_0}, \ldots, \xi_{\nu_3}$, the standardised maximin c-optimal design ξ_{ν_4} and the two-point cluster design ξ_1.

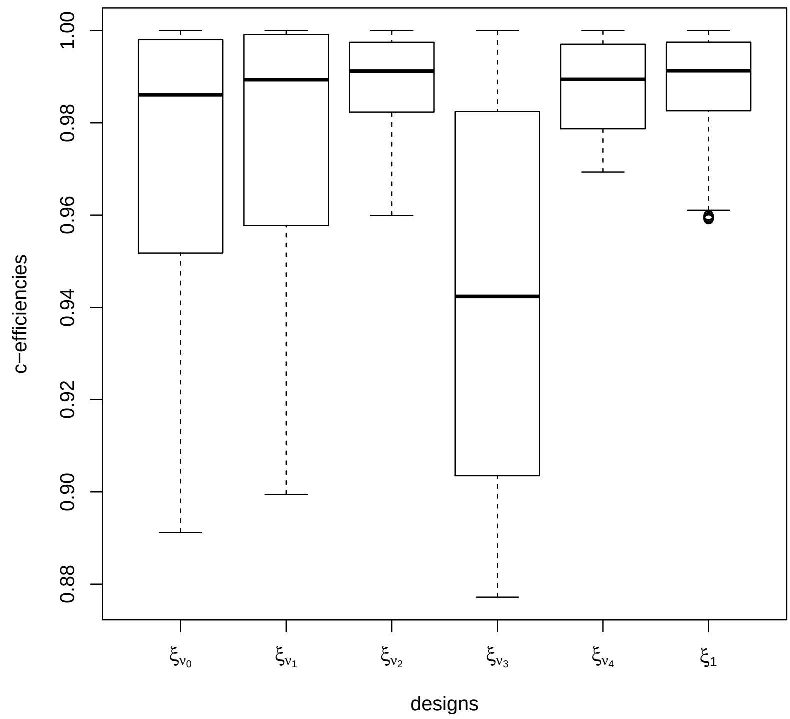


Figure 5.2: Boxplots of c-efficiencies calculated for 6 designs for 1000 parameter vectors

Among the locally c-optimal designs $\xi_{\nu_0}, \ldots, \xi_{\nu_3}$, only ξ_{ν_2} performs well across the assumed parameter space while the rest are not so good for some scenarios. As for the D-optimality case, the standardised maximin c-optimal design ξ_{ν_4} has the highest minimum but not the best median efficiency amongst the designs. Hence there is a trade off between seeking a high minimum efficiency and a high median efficiency. Either of the two-point cluster design ξ_1 or the standardised maximin c-optimal design ξ_{ν_4} can be used instead of the locally c-optimal designs to achieve robustness.

Chapter 6

Optimal designs for partial likelihood information

For a model under the proportional hazards parametrisation (see, for example, equation (4.2)), often interest centres on estimating just the coefficients of the covariates which represent the explanatory variable effects; the baseline hazard and any unknown parameters involved in it are of secondary interest (see the motivating example discussed in Chapter 5). Therefore, here we consider the construction of optimal designs for Cox's proportional hazards model, introduced by Cox (1972), that leaves the baseline hazard function unspecified. The main assumption of this model, also known as the Cox regression model, is that of proportional hazards over time. This means that the hazard function for any subject under investigation is multiplicatively related to the hazard function of another subject, that is, their hazard ratio is constant over time.

When Cox's model is considered the full likelihood function cannot be used for the estimation of the vector of covariate parameters since the baseline hazard function is of arbitrary form. Hence the designs found so far are not directly applicable here. However, inference on the explanatory variables coefficients can be made based on the partial likelihood function proposed by Cox (1972) which does not require knowledge of the baseline hazard. Therefore, for the construction of optimal designs, the asymptotic covariance matrix of Cox's partial, rather than full, likelihood estimator for the covariate coefficients is considered.

Andersen and Gill (1982) formulate Cox's model in a counting process set-up and provide analytical results for the asymptotic properties of the estimators from this model. However, there are only two papers in the literature so far that consider optimal designs for Cox's model. Kalish and Harrington (1988) find optimal designs for the special case when two treatments are available and investigate empirically how much efficiency is lost when equal numbers of patients are allocated to the two treatments, that is, when the standard design is used. López-Fidalgo and Rivas-López (2012) derive a partial information matrix for the covariate coefficients using approximations, as well as the information matrix for the full likelihood model. In their application, they also consider a binary design space and find optimal designs for the partial likelihood model which they then compare with the optimal designs for the full likelihood model.

This chapter is organised as follows. First Cox's proportional hazards model involving a vector of explanatory variables and the corresponding partial likelihood function are presented. We also find a general expression for the asymptotic covariance matrix of Cox's partial likelihood estimator for the covariate coefficients. Our approach is then illustrated through an application to the special case of only one covariate for which we derive the optimality criterion to be used and find a necessary condition for the optimality of a design. Minimum variance designs are computed for different censoring mechanisms and for binary and interval design spaces. Finally, we compare these designs with the ones found in Chapter 4 for the corresponding parametric model that involves a full likelihood information and demonstrate that the latter designs are highly efficient for estimation in the partial likelihood model.

6.1 Cox's model and partial likelihood function

When the risk of the desired event occurring at a particular time t depends on the values of a set of explanatory variables, Cox's proportional hazards model is specified by the hazard function

$$h(t, x_j) = h_0(t)e^{\boldsymbol{\beta}^\top \boldsymbol{x}_j} \quad t > 0, \tag{6.1}$$

where $\boldsymbol{x}_j$ is the value of the covariate vector for the jth subject in the experiment and $\boldsymbol{\beta}$ is the vector of coefficients of the explanatory variables in the model that needs to be estimated. The function $h_0(t)$ is the baseline hazard function which remains unspecified and can be regarded as the hazard function for a subject for which the values of all the explanatory variables are zero.

The linear combination $\boldsymbol{\beta}^\top \boldsymbol{x}_j$ of the explanatory variables in $\boldsymbol{x}_j$ is called the linear component of the model and does not include a constant term since in such a case, it would cancel out by a simple rescaling of $h_0(t)$. Furthermore, the relative hazard $e^{\boldsymbol{\beta}^\top \boldsymbol{x}_j}$, also known as the hazard ratio, is the ratio of the hazard at time t for the jth subject with covariate values vector $\boldsymbol{x}_j$ relative to the hazard for a subject with $\boldsymbol{x}_j = \mathbf{0}$. Therefore, each parameter in the coefficient vector $\boldsymbol{\beta}$ explains how the hazard changes with respect to the corresponding explanatory variable. In particular, if a β-parameter is positive then the risk of the event of interest occurring increases with that covariate, whereas negative values correspond to the explanatory variable having a decreasing effect on the hazard.

The hazard ratio is always non-negative with $e^{\boldsymbol{\beta}^\top \boldsymbol{x}_j}$ being the most commonly used choice for specifying it (see Collett (2003)). Finally, we note that the hazard ratio is independent of t and so model (6.1) satisfies the proportional hazards assumption of constant hazard ratio over time. However, the baseline hazard function, and hence the probability distribution of the times to event, is not specified and therefore Cox's proportional hazards model is referred to as a semi-parametric model. It is this flexibility of the Cox model, together with the simple interpretability of the regression coefficients in terms of hazard ratios, that has made the model so popular in survival studies; see, for example, Collett (2003) for details and examples.

Cox (1972) shows that under the proportional hazards assumption the β-parameters can be estimated without making any further assumptions on the particular form of the distribution for the times to event. He proposes the use of a conditional likelihood, referred to as the partial likelihood that enables this estimation.

Suppose that data are available for n subjects with corresponding observations denoted by $y_1, \ldots, y_n$ and that δ_j is an indicator function which is equal to zero if the jth observation, y_j, $j = 1, \ldots, n$ is right-censored and unity otherwise. The partial likelihood function for model (6.1) is

$$L_P(\boldsymbol{\beta}) = \prod_{j=1}^{n} \left\{ \frac{e^{\boldsymbol{\beta}^\top \boldsymbol{x}_j}}{\sum_{l \in R(y_j)} e^{\boldsymbol{\beta}^\top \boldsymbol{x}_l}} \right\}^{\delta_j}, \tag{6.2}$$

with corresponding log-likelihood function given by

$$\log L_P(\boldsymbol{\beta}) = \sum_{j=1}^{n} \delta_j \left\{ \boldsymbol{\beta}^\top \boldsymbol{x}_j - \log \sum_{l \in R(t_j)} e^{\boldsymbol{\beta}^\top \boldsymbol{x}_l} \right\}. \tag{6.3}$$

The set $R(y_j)$ is called the risk-set at time y_j and contains the indices of those subjects for which neither the event of interest nor censoring have occurred at a time just prior to y_j. For example, if the event under investigation is death then $R(y_j)$ is the set of indices of patients who are alive and their corresponding observations are uncensored at a time just before y_j. The subjects with indices included in the risk-set $R(y_j)$ are said to be at risk at time y_j.

Note that any $1-1$ increasing transformation of the y_j leaves (6.2) unchanged. Therefore, the log-likelihood function in (6.3) and hence inference on the β-parameters, depend only on the order of occurrence of the observed event times.

Assuming there are no ties in the data, the partial likelihood defined in (6.2) can be obtained as a conditional probability conditioning on the observed event times (see section 2.2.3 in Chapter 2 for the derivation). The actual censored and uncensored times to event are not used directly and so this is not a true likelihood. A brief discussion on the treatment of ties is also given in section 2.2.3 of Chapter 2.

6.2 Optimality criterion

Let $\hat{\boldsymbol{\beta}}_{PL}$ be the maximum partial likelihood estimator of the explanatory variables coefficients vector $\boldsymbol{\beta}$. This estimator is defined as the solution of the likelihood equation

$$\frac{\partial \log L_P(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0,$$

where $\log L_P(\boldsymbol{\beta})$ is given in equation (6.3), with its asymptotic variance determined by the inverse of

$$E\left[-\frac{\partial^2 \log L_P(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right].$$

López-Fidalgo and Rivas-López (2012) approximate this expectation for one covariate and find the optimal designs by maximising the resulting expression. They therefore add an extra layer of approximation to the optimality criterion, in addition to the fact that the information matrix in itself approximates the inverse of the covariance matrix. For large sample sizes, the latter approximation converges. However we have found that the quality of the former approximation is questionable by comparing their results on the information matrix to the asymptotic variance matrix provided by Andersen and Gill (1982).

We work directly with the asymptotic covariance matrix which we derive from Andersen and Gill (1982) who prove that under some asymptotic regularity conditions

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{PL} - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma^{-1}), \tag{6.4}$$

where $\mathbf{0}$ is the zero vector of appropriate length. Now let ξ be an approximate design with support points x_i, $i = 1, \ldots, m$ and corresponding weights ω_i, $i = 1, \ldots, m$. For model (6.1) the inverse, $\Sigma = \Sigma(\xi)$, of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_{PL}$ is given by

$$\Sigma = \Sigma(\xi) = \sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q e^{\boldsymbol{\beta}^\top (\boldsymbol{x}_i + \boldsymbol{x}_q)} (\boldsymbol{x}_i - \boldsymbol{x}_q)(\boldsymbol{x}_i - \boldsymbol{x}_q)^\top \int_0^\infty \frac{\pi_i(y)\pi_q(y)h_0(y)}{\sum_{l=1}^{m} \omega_l \pi_l(y) e^{\boldsymbol{\beta}^\top \boldsymbol{x}_l}} \, dy, \tag{6.5}$$

where $\pi_i(y)$, $i = 1, \ldots, m$ is the probability that a subject with covariate values vector $\boldsymbol{x}_i$ is at risk at time y, that is, neither the event nor censoring have occurred by time y. The "risk function" $\pi_i(y)$ differs according to the censoring scheme considered.

An optimal design for model (6.1) minimises the asymptotic covariance matrix of the maximum partial likelihood estimate $\hat{\beta}_{PL}$ or equivalently maximises $\Sigma(\xi)$, with respect to the design ξ. Thus a design ξ^* is optimal for estimating $\boldsymbol{\beta}$ if

$$\xi^* = \arg\min_{\xi} \Sigma^{-1}(\xi) = \arg\max_{\xi} \Sigma(\xi),$$

where $\Sigma(\xi)$ is given in (6.5). We note that the optimal design will depend on the values of the β-parameters and therefore will be a locally optimal design.

It is clear from the asymptotic distribution given in (6.4) that the bias of the estimator $\hat{\boldsymbol{\beta}}_{PL}$ is of order $o(n^{-1/2})$. Hence the variance will dominate the mean squared error for large n, thus justifying our choice of optimality criterion, which is solely based on the asymptotic covariance matrix.

For illustration purposes, in what is to follow, we will consider the special case of only one covariate being involved in Cox's proportional hazards model and therefore there is only one β-parameter that requires estimation. This situation is often encountered in clinical trials where patients are randomised to different treatments or doses of a treatment. Similarly, in accelerated life testing, there is usually just one covariate to be selected by the experimenter.

Proposition 1 gives a necessary condition for the optimality of a design ξ^*, that is, a design that does not satisfy this condition cannot be optimal. Its proof is given in appendix B, section B.2.1. Unlike the equivalence theorem for c-optimality, this condition is not sufficient, since the criterion function, $\Sigma(\xi)$, could not be shown to be concave.

Proposition 1. *Let $\mathcal{H}$ be the class of all one-point designs where the support point is in the design space $\mathcal{X} = [u, v]$, and let $\eta = \{x; 1\} \in \mathcal{H}$. If a design ξ^* on $\mathcal{X}$ with support points $\{x_1, \ldots, x_m\}$ and corresponding weights $\{\omega_1, \ldots, \omega_m\}$ is optimal for estimating β via the partial likelihood method, the inequality*

$$d(\xi^*, \eta) \leq 0$$

holds for all $\eta \in \mathcal{H}$, with equality in the one-point designs $\xi_i = \{x_i; 1\}$, $i = 1, \ldots, m$, generated by the support points of ξ^*. Here $d(\xi^*, \eta)$ is the Frechet derivative of the criterion function at ξ^* in direction of the one-point design η, and is given by

$$\begin{aligned} d(\xi^*, \eta) \quad = \quad & -\sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q e^{\beta(x_i + x_q)} (x_i - x_q)^2 \int_0^\infty \frac{h_0(y) \pi_i(y) \pi_q(y)}{\sum_{l=1}^{m} \omega_l \pi_l(y) e^{\beta x_l}} dy \\ & -\sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q e^{\beta(x_i + x_q)} (x_i - x_q)^2 \int_0^\infty \frac{h_0(y) \pi_i(y) \pi_q(y) \pi_x(y) e^{\beta x}}{(\sum_{l=1}^{m} \omega_l \pi_l(y) e^{\beta x_l})^2} dy \\ & + \sum_{q=1}^{m} \omega_q e^{\beta(x + x_q)} (x - x_q)^2 \int_0^\infty \frac{h_0(y) \pi_x(y) \pi_q(y)}{\sum_{l=1}^{m} \omega_l \pi_l(y) e^{\beta x_l}} dy, \end{aligned}$$

where $\pi_x(y)$ is the probability of being at risk at time y given covariate value x.

In the case of a binary design space $\mathcal{X} = \{0, 1\}$, the design problem is that of allocating the subjects, for example, to two treatments and so the design must be supported at points $x_1 = 0$ and $x_2 = 1$. Let ω, $1 - \omega$ be the weights at points 0 and 1 respectively. Using the results of Andersen and Gill (1982), Kalish and Harrington (1988) find the asymptotic variance of $\sqrt{n}(\hat{\beta}_{PL} - \beta)$ to be

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta}} \left[\int_0^\infty \frac{\pi_1(y)\pi_2(y)h_0(y)}{\omega\pi_1(y) + (1-\omega)e^{\beta}\pi_2(y)} dy \right]^{-1}, \qquad (6.6)$$

where $\pi_i(y), i = 1, 2$ is the probability of being at risk at time y for subjects allocated to points 0 and 1 respectively.

For purposes of comparison with the c-optimal designs found in Chapter 4, for a continuous design space we consider designs with support points x_1 and x_2 which represent, for example, drug doses and corresponding weights ω and $1 - \omega$. From (6.5) the asymptotic variance of $\sqrt{n}(\hat{\beta}_{PL} - \beta)$ can be written as

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2 - x_1)^2} \left[\int_0^\infty \frac{\pi_1(y)\pi_2(y)h_0(y)}{\omega e^{\beta x_1}\pi_1(y) + (1-\omega)e^{\beta x_2}\pi_2(y)} dy \right]^{-1}, \quad (6.7)$$

where $\pi_1(y)$ ($\pi_2(y)$) is the "risk function" for subjects allocated to dose x_1 (x_2).

The optimal design in each case is found by minimising $\Sigma^{-1}(\xi)$ or, equivalently, maximising $\Sigma(\xi)$ with respect to the design ξ. We note that all optimal designs we found for continuous design intervals $\mathcal{X}$ are supported at exactly two points. Therefore the criterion in (6.7) is given for two-point designs.

6.3 Minimum variance designs using the partial likelihood method

In this section we present the optimal designs that minimise criteria (6.6) and (6.7), for a binary and a continuous design space respectively, assuming a constant baseline hazard function. This corresponds to the most widely used proportional hazards model based on the exponential distribution.

We first discuss the special case of no censoring and then consider both Type-I and random censoring mechanisms. The designs are found numerically for various β-values and percentages of censoring and our results are compared with those of Kalish and Harrington (1988). We further extend a result by Kalish and Harrington (1988) to interval design spaces, where we show that for Type-I censoring the optimal designs do not depend on the shape of the baseline hazard function. Therefore the designs found for the exponential-based proportional hazards model are applicable to model (6.1) with one covariate for any form of $h_0(t)$.

Let $Y_1, \ldots, Y_n$ be independent random variables for the times to event of the n subjects in the experiment with $y_1, \ldots, y_n$ their corresponding observed values and

$[0, c]$ be the period of the experiment. Following Kalish and Harrington (1988), the survivor function of the random variable W representing the time to censoring is given by

$$S_W(w) = \begin{cases} 1, & \text{if } 0 < w \leq c \\ 0, & \text{if } w > c \end{cases}, \qquad S_W(w) = \begin{cases} \frac{c-w}{c}, & \text{if } 0 < w \leq c \\ 0, & \text{if } w > c \end{cases}$$

for Type-I and random censoring respectively. Therefore, the probability that a subject allocated to x_i is at risk at time y is $\pi_i(y) = S_W(w)S_i(y)$, $i = 1, 2$, where $S_i(y)$ is the survivor function of the times to event for subjects allocated to x_i.

We also use the Kalish and Harrington (1988) characterisation for the 'amount of censoring' as the overall probability of censoring if a balanced design with equal weights at the two support points had been used. That is $1 - (0.5d_1 + 0.5d_2)$, where $d_i = P(Y_j < W) = \int_0^\infty S_W(y)dF_i(y)$ is the probability of the event occurring and $F_i(y)$ is the distribution function of the times to event, for subjects allocated to x_i, $i = 1, 2$.

6.3.1 No censoring

The no censoring case corresponds to $c = \infty$, that is, an experiment that runs for as long as necessary to record all times until the occurrence of the event of interest. In this case $\pi_i(y) = S_i(y)$, $i = 0, 1$ and equations (6.6) and (6.7) can be written as

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^\beta} \left[\int_0^1 \frac{u^{e^\beta - 1}}{\omega + (1-\omega)e^\beta u^{e^\beta - 1}} du \right]^{-1} \tag{6.8}$$

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2 - x_1)^2} \left[\int_0^1 \frac{u^{e^{\beta x_2} - 1}}{\omega e^{\beta x_1} + (1-\omega)e^{\beta x_2} u^{e^{\beta x_2} - e^{\beta x_1}}} du \right]^{-1}, \tag{6.9}$$

respectively using the parametrisation $u = S_0(y) = \exp\{-\int_0^y h_0(s)ds\}$. Then $S_i(y) = \{S_0(y)\}^{e^{\beta x_i}}$, $i = 1, 2$ and $\lim_{y\to\infty} S_0(y) = 0$ since all times until the occurrence of the event are observed. Therefore, whether a binary or a continuous design space is considered, the baseline hazard does not affect the optimal choice of design.

Assuming exponential times to event, the optimal designs on the binary and the continuous design space were constructed for various β-values using the Gauss-Laguerre approximation to the integrals involved in (6.8) and (6.9) respectively and then minimising the resulting expressions. These are presented in Table 6.1 where $1 - \omega$ is the optimal weight corresponding to the larger support point of the design x_2. We note that the continuous design interval considered in these calculations is $\mathcal{X} = [0, 1]$. For every choice of β, the efficiency of the balanced design is also found as

$$eff(\xi_{bal}) = \frac{\Sigma^{-1}(\xi_{loc})}{\Sigma^{-1}(\xi_{bal})},$$

where ξ_{loc} is the locally optimal design corresponding to that β-value and ξ_{bal} is the balanced design allocating half the observations at 0 and the rest at point 1.

Table 6.1: Optimal designs for binary and continuous design spaces and efficiencies, in percent, of the balanced design in the absence of censoring

optimal	$e^{\beta}(\beta)$					
design	0.03 (-3.51)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	33.3 (3.51)
$1-\omega$	0.68	0.55	0.51	0.49	0.45	0.32
efficiency	(92)	(99)	(100)	(100)	(99)	(91)
$\{x_1, x_2\}$	{0.04,0.96}	{0,1}	{0,1}	{0,1}	{0,1}	{0.1,1}
$1-\omega$	0.66	0.55	0.51	0.49	0.45	0.34
efficiency	(90)	(99)	(100)	(100)	(99)	(90)

From Table 6.1 we observe that for a positive value of β the optimal weight $1-\omega$ at point $x_2 = 1$ is the same as the weight ω at point $x_1 = 0$ for the corresponding negative β of equal absolute value. Moreover, for small and moderate absolute values of β, that is 0.69 and 1.39 the efficiency of the balanced design is very high and decreases for larger absolute values of β ($|\beta| = 3.51$). Using the maximum likelihood estimate for β found for the Freireich data, see Freireich et al. (1963), given by $\hat{\beta} = -1.526$ the optimal design is always supported at points 0 and 1 allocating a proportion of 0.56 subjects at point 1 for both a binary and a continuous design space. The efficiency of the balanced design is found in both cases to be 98%

6.3.2 Type-I censoring

Under Type-I censoring Kalish and Harrington (1988) showed that equation (6.6) can be written as

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta}} \left[\int_{S_0(c)}^{1} \frac{u^{e^{\beta}-1}}{\omega + (1-\omega)e^{\beta}u^{e^{\beta}-1}} \, du \right]^{-1}, \tag{6.10}$$

where $S_0(y) = exp\{-\int_0^y h_0(s)ds\}$. We extend this result to the case of a continuous design space. Using the fact that under proportional hazards $S_i(y) = \{S_0(y)\}^{e^{\beta x_i}}$, $i = 1, 2$ and applying the transformation $u = S_0(y)$ equation (6.7) becomes

$$\begin{aligned}\Sigma^{-1}(\xi) &= \frac{1}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2-x_1)^2} \left[\int_0^c \frac{S_1(y)S_2(y)h_0(y)}{\omega e^{\beta x_1}S_1(y) + (1-\omega)e^{\beta x_2}S_2(y)} \, dy \right]^{-1} \\ &= \frac{1}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2-x_1)^2} \left[\int_1^{S_0(c)} \frac{-u^{e^{\beta x_1}}u^{e^{\beta x_2}}}{\omega e^{\beta x_1}u^{e^{\beta x_1}} + (1-\omega)e^{\beta x_2}u^{e^{\beta x_2}}} \, du \right]^{-1}\end{aligned}$$

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)e^{\beta(x_1+x_2)}(x_2-x_1)^2}\left[\int_{S_0(c)}^{1} \frac{u^{e^{\beta x_2}-e^{\beta x_1}}}{(1-\omega)e^{\beta x_1}+\omega e^{\beta x_2}u^{e^{\beta x_2}-e^{\beta x_1}}}du\right]^{-1}. \tag{6.11}$$

In both cases $\Sigma^{-1}(\xi)$ depends on the baseline hazard only through $S_0(c)$ for which we can assume that a good approximation of its value is available by the experimenter. In particular, $S_0(c)$ is the probability of the event occurring at a time equal or greater than c for patients allocated at point $x = 0$ which corresponds, for example, to a placebo or a standard treatment previously investigated in depth.

In conclusion, under Type-I censoring and for both binary and continuous design spaces, the optimal design is independent of the shape of $h_0(y)$ and for its construction we can assume without loss of generality a constant baseline hazard corresponding to the exponential-based proportional hazards model. Therefore, the designs for the exponential regression model are optimal for partial likelihood estimation for all proportional hazards models.

Table 6.2 shows the optimal weight $1-\omega$ at the larger support point $x_2 = 1$ of the design on $\mathcal{X} = \{0, 1\}$ and the efficiency of the balanced design assuming the exponential regression model. For various values of the covariate hazard ratio e^β and the amount of censoring, an approximation to the integral given in (6.10) is made using the Gauss-Legendre method, followed by a minimisation of the resulting expression for $\Sigma^{-1}(\xi)$.

The choice of the absolute β-values presented in Table 6.2 was made so that we account for small, moderate as well as large treatments effects. We also include the value 0.3 of the amount of censoring as this was used in the Freireich data. A percentage of censoring as high as 90% can occur in reliability studies where, for example, a component of a machine is proven to be very reliable.

Table 6.2: Optimal weights $1-\omega$ corresponding to $x_2 = 1$ and efficiency, in percent, of the balanced design for a binary design space and Type-I censoring

amount of censoring	$e^\beta(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	0.68 (92)	0.60 (97)	0.55 (99)	0.52 (100)	0.48 (100)	0.45 (99)	0.40 (97)	0.32 (92)
0.3	0.68 (92)	0.61 (96)	0.58 (98)	0.54 (99)	0.46 (99)	0.42 (98)	0.39 (96)	0.32 (92)
0.5	0.76 (80)	0.68 (88)	0.62 (95)	0.56 (99)	0.44 (99)	0.38 (95)	0.32 (88)	0.24 (80)
0.7	0.82 (71)	0.73 (83)	0.64 (93)	0.57 (98)	0.43 (98)	0.36 (93)	0.27 (83)	0.18 (71)
0.9	0.85 (68)	0.75 (80)	0.66 (91)	0.58 (97)	0.42 (97)	0.34 (91)	0.25 (80)	0.16 (68)

We observe that for values of e^β greater than unity, that is, for positive β-values, the weight $1-\omega$ is smaller than 0.5 whereas for $e^\beta < 1$, that is $\beta < 0$, the optimal design allocates more subjects to point $x_2 = 1$. This agrees with our results on locally c-optimal designs arising from the full likelihood method discussed in section 4.4.4. Under both methods the optimal design allocates more subjects to the experimental point where the possibility of censoring is greater in order for the variance to be minimised. This point is $x_1 = 0$ when the parameter β is positive since in this case the probability of occurrence of the event of interest is larger at point 1. Moreover, as for the no censoring case we can observe the symmetry in the optimal weights ω and $1-\omega$ at points 0 and 1 respectively for the same absolute β-values.

From Table 6.2 it is also evident that for heavy censoring (above 50%) and absolute β-values moderately away from zero ($|\beta| \geq 2.3$) the efficiency of the balanced design, with equally supported points 0 and 1, relative to the optimal design, drops below 90%. This contradicts the results by Kalish and Harrington (1988) who only look at small values for the parameter β and therefore find the balanced design to be very efficient.

We now consider the continuous design space $\mathcal{X} = [0, 1]$. The support points $\{x_1, x_2\}$ of the optimal design, the optimal weight $1-\omega$ corresponding to the larger support point x_2 and the efficiencies of the balanced design are presented in Table 6.3 for the same values of β and amount of censoring used for the binary design space case. The Gauss-Legendre approximation to the integral in (6.11) is again used before the minimisation of $\Sigma^{-1}(\xi)$.

Table 6.3: Support points $\{x_1, x_2\}$, optimal weights $1-\omega$ at point x_2 and efficiency, in percent, of the balanced design under Type-I censoring for $\mathcal{X} = [0, 1]$

amount of censoring	$e^\beta(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	{0.04,0.96} 0.66 (90)	{0,1} 0.60 (97)	{0,1} 0.55 (99)	{0,1} 0.52 (100)	{0,1} 0.48 (100)	{0,1} 0.45 (99)	{0,1} 0.40 (97)	{0.04,0.96} 0.34 (90)
0.3	{0,0.91} 0.66 (90)	{0,1} 0.61 (96)	{0,1} 0.58 (98)	{0,1} 0.54 (99)	{0,1} 0.46 (99)	{0,1} 0.42 (98)	{0,1} 0.39 (96)	{0.09,1} 0.34 (90)
0.5	{0,0.84} 0.71 (76)	{0,1} 0.68 (88)	{0,1} 0.62 (95)	{0,1} 0.56 (99)	{0,1} 0.44 (99)	{0,1} 0.38 (95)	{0,1} 0.32 (88)	{0.16,1} 0.29 (76)
0.7	{0,0.77} 0.76 (63)	{0,1} 0.73 (83)	{0,1} 0.64 (93)	{0,1} 0.57 (98)	{0,1} 0.43 (99)	{0,1} 0.36 (93)	{0,1} 0.27 (83)	{0.23,1} 0.24 (63)
0.8	{0,0.74} 0.78 (59)	{0,1} 0.75 (80)	{0,1} 0.66 (91)	{0,1} 0.58 (97)	{0,1} 0.42 (97)	{0,1} 0.34 (91)	{0,1} 0.25 (80)	{0.26,1} 0.22 (59)

For large absolute β-values and heavy censoring the support points of the optimal design move away from the end-points 0 and 1 of the design space. In particular, the optimal design always includes the most informative points about the probability of occurrence of the event of interest, that is, the points where censoring is less likely. These will be experimental points greater than zero for $e^{\beta} > 1$ in which case the covariate has an increasing effect on the hazard, and points smaller than unity when $e^{\beta} < 1$, that is $\beta < 0$, and the probability of the event occurring decreases with the explanatory variable.

The symmetry for the optimal weights for some absolute values of β is evident as well as the symmetry of the support points. This is, for large positive values of β the smaller support point of the design moves away from the boundary 0 by the same amount the larger support point is away from 1 for the corresponding negative β of the same absolute value. In the case of the amount of censoring being equal to 0.1, that is, 10% of the observations will be right-censored, and absolute β-value equal to 3.51 the optimal design is not supported at either of the boundaries 0 and 1 of the design space.

We also note that for small and moderate values of the parameter β the design is supported at points 0 and 1. However, in every case the weights of the optimal design follow the same pattern as the one discussed above for a binary design space, thus making the equal allocating balanced design suboptimal. Moreover, for absolute β-values greater than 2.3 and heavy censoring, above 50%, the efficiency of the balanced design drops substantially below 90%.

6.3.3 Random censoring

In the presence of random censoring the criteria functions $\Sigma^{-1}(\xi)$ for binary and continuous design spaces are given by

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)ce^{\beta}} \left[\int_0^c \frac{(c-y)S_1(y)S_2(y)h_0(y)}{\omega S_1(y) + (1-\omega)S_2(y)}\, dy \right]^{-1}, \tag{6.12}$$

and

$$\Sigma^{-1}(\xi) = \frac{1}{\omega(1-\omega)ce^{\beta(x_1+x_2)}(x_2-x_1)^2} \left[\int_0^c \frac{(c-y)S_1(y)S_2(y)h_0(y)}{\omega e^{\beta x_1} S_1(y) + (1-\omega)e^{\beta x_2} S_2(y)}\, dy \right]^{-1}, \tag{6.13}$$

respectively.

A similar transformation as the one used for Type-I censoring can not be applied here and therefore $\Sigma^{-1}(\xi)$ and hence the optimal design does depend on the form of the underlying hazard. An explanation for this is that for Type-I censoring a $1-1$ increasing transformation of the exponentially distributed times will not affect the partial likelihood function and so $\Sigma^{-1}(\xi)$ will remain the same. However, under random censoring transforming the assumed uniform distribution for the censoring times will result in the distribution no longer being uniform. Therefore the same generalisation does not hold under the random censoring scenario.

For illustration purposes we compute the optimal designs for various β-values and amounts of censoring again assuming a constant baseline hazard and applying the Gauss-Legendre approximation to the integrals given in (6.12) and (6.13). These designs are displayed in Tables 6.4 and 6.5 for the cases of a binary, $\mathcal{X} = \{0, 1\}$, and a continuous, $\mathcal{X} = [0, 1]$, design space respectively along with the efficiencies of the balanced design.

Table 6.4: Optimal weights $1 - \omega$ corresponding to $x_2 = 1$ and efficiency, in percent, of the balanced design for a binary design space and random censoring

amount of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	0.68 (91)	0.61 (97)	0.55 (99)	0.52 (100)	0.48 (100)	0.45 (99)	0.39 (97)	0.32 (91)
0.3	0.68 (91)	0.62 (96)	0.57 (98)	0.53 (100)	0.47 (100)	0.43 (98)	0.38 (96)	0.32 (91)
0.5	0.71 (87)	0.65 (92)	0.60 (96)	0.55 (99)	0.45 (99)	0.40 (96)	0.35 (92)	0.94 (87)
0.7	0.81 (73)	0.71 (85)	0.63 (94)	0.57 (98)	0.43 (98)	0.37 (94)	0.29 (85)	0.19 (73)
0.9	0.84 (68)	0.75 (80)	0.66 (91)	0.58 (97)	0.42 (97)	0.34 (91)	0.25 (80)	0.16 (68)

As before, we observe that the optimal design is supported at the experimental points where the probability of censoring and therefore the variance is smaller, and puts more weight at the support point where censoring is more likely so that the variance at that point is minimised.

For negative and positive β's of the same absolute value, the smaller (larger) support point of the design moves away from the 0 (1) boundary of the design space by the same amount. The optimal weights at points 0 and 1 are also equal for negative and positive β-values of the same absolute value.

Overall the two censoring schemes produce similar designs which differ from the balanced design for heavy censoring and absolute β-values moderately far from the trivial case of $\beta = 0$. In particular, for $|\beta| \geq 2.3$ and more than 50% right-censored observations the balanced design has efficiencies less than 90%.

Table 6.5: Support points $\{x_1, x_2\}$, optimal weights $1-\omega$ at point x_2 and efficiency, in percent, of the balanced design under random censoring for $\mathcal{X} = [0, 1]$

amount of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	{0,0.91} 0.66 (90)	{0,1} 0.61 (97)	{0,1} 0.55 (99)	{0,1} 0.52 (100)	{0,1} 0.48 (100)	{0,1} 0.45 (99)	{0,1} 0.39 (97)	{0.09,1} 0.34 (90)
0.3	{0,0.91} 0.66 (90)	{0,1} 0.62 (96)	{0,1} 0.57 (98)	{0,1} 0.53 (100)	{0,1} 0.47 (100)	{0,1} 0.43 (98)	{0,1} 0.38 (96)	{0.09,1} 0.34 (90)
0.5	{0,0.88} 0.68 (85)	{0,1} 0.65 (92)	{0,1} 0.60 (96)	{0,1} 0.55 (99)	{0,1} 0.45 (99)	{0,1} 0.40 (96)	{0,1} 0.35 (92)	{0.12,1} 0.32 (85)
0.7	{0,0.79} 0.75 (67)	{0,1} 0.71 (85)	{0,1} 0.63 (94)	{0,1} 0.57 (98)	{0,1} 0.43 (98)	{0,1} 0.37 (94)	{0,1} 0.29 (85)	{0.21,1} 0.25 (67)
0.8	{0,0.74} 0.77 (60)	{0,1} 0.75 (80)	{0,1} 0.66 (91)	{0,1} 0.58 (97)	{0,1} 0.42 (97)	{0,1} 0.34 (91)	{0,1} 0.25 (80)	{0.26,1} 0.23 (60)

6.4 Comparison of designs arising from full and partial likelihood methods

Efron (1977) compares the Fisher information for estimating β, for both the full and the partial likelihood methods, in the same underlying model. He finds that the Fisher informations coincide except for an extra term in the Fisher information for the full likelihood, which, however, will usually be small in practice. He concludes that therefore in most situations the partial likelihood method will be reasonably efficient. A simple explicit formula for this extra term could not be derived even in the simple case of a binary design space. Therefore, we could not work directly with the Fisher information matrix to prove the similarity of the two approaches analytically.

However, these results suggest that the optimal designs for estimating β, which are based on the asymptotic variances and thus the Fisher information, should also be similar. In particular, we wish to find out in which situations the optimal designs for the full likelihood method, which are constructed in Chapter 4, are efficient for estimation in the partial likelihood model. Hence finding optimal designs for the complicated criterion function $\Sigma(\xi)$ could be avoided by practitioners.

We first compare the optimal designs for several scenarios directly, to see in which situations they are similar or even coincide and then find an explanation for this phenomenon.

Throughout this section, we assume an exponential regression model with constant baseline hazard $h_0(t) \equiv \exp(\alpha)$ for some constant $\alpha \in I\!R$. Then the hazard function is $h(t, x) = h_0(t) \exp(\beta x) = \exp(\alpha + \beta x)$, and we compare the locally c-optimal design for estimating β in the two-parameter model with the Σ-optimal design for β in Cox's model.

We note that López-Fidalgo and Rivas-López (2012) provide a brief comparison of such designs for a binary design space. However, they assume that $\exp(\alpha) = 1$, leaving them with an estimation problem for one parameter only. Hence the optimal designs they find for the parametric model are one-point designs, taking all observations at $x = 1$. This is not surprising since they completely specify the baseline hazard, implying that the hazard at $x = 0$ is known, thus not requiring any observations at $x = 0$.

6.4.1 Numerical results

We briefly discuss the case of no censoring for which the locally c-optimal design for β found using the full likelihood function is always equally supported at 0 and 1 for both a binary and a continuous design space (see section 4.4.2). From Table 6.1 we observe that for β-values away from zero the two approaches do not coincide as the optimal weights for the partial likelihood method are not equal. However, the balanced design allocating equal number of subjects at points 0 and 1 is highly efficient even for large values of the parameter β making the locally c-optimal designs for β good alternatives to the designs found using the partial likelihood function.

In the presence of censoring, we calculate the efficiency of the locally c-optimal designs found using the full likelihood function relative to the designs discussed in sections 6.3.2 and 6.3.3 by

$$eff(\xi_F^*) = \frac{\Sigma(\xi_F^*)}{\Sigma(\xi_P^*)} = \frac{\Sigma^{-1}(\xi_P^*)}{\Sigma^{-1}(\xi_F^*)},$$

where ξ_F^* and ξ_P^* are the locally optimal designs for β arising from the full and partial likelihood method respectively. The results for the two censoring schemes considered are illustrated in Tables 6.6 and 6.7 respectively. The cases of $\mathcal{X} = \{0, 1\}$ and $\mathcal{X} = [0, 1]$ are examined and the efficiencies are found as functions of the amount of censoring and the parameter of interest β.

We observe that the c-optimal designs found using the full likelihood function are extremely efficient under both censoring schemes, with the efficiencies under random censoring being slightly lower. Hence the c-optimal designs can be used as an efficient alternative for the Σ-optimal designs, even if the data are to be analysed through the partial likelihood approach. In particular, for heavy censoring the full likelihood designs give efficiency very close or equal to 1 even for extremely large β-values.

Table 6.6: Efficiencies, in percent, of full likelihood designs under Type-I censoring for a binary (and a continuous) design space

amount of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	94 (93)	98 (98)	100 (100)	100 (100)	100 (100)	100 (100)	98 (98)	94 (93)
0.3	99 (98)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	99 (98)
0.5	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
0.7	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
0.9	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

Table 6.7: Efficiencies, in percent, of full likelihood designs under random censoring for a binary (and a continuous) design space

amount of censoring	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	94 (92)	98 (98)	100 (100)	100 (100)	100 (100)	100 (100)	98 (98)	94 (92)
0.3	98 (97)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	98 (97)
0.5	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
0.7	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
0.9	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

Comparing the elements of Tables 6.6 and 6.7, that is, the efficiencies of the c-optimal designs, with the corresponding elements in Tables 6.2-6.5 we find that the c-optimal designs are considerably more efficient for estimating β in the partial likelihood model than the balanced design on 0 and 1. For example, when the percentage of censoring is 0.5 and $\beta = -2.3$, the c-optimal designs have efficiencies of 100% for Type-I and random censoring, respectively for both design spaces whereas the balanced design achieves corresponding efficiencies of 88% for Type-I censoring and 93% for random censoring again for both design spaces. This means that under Type-I censoring we require 114 individuals in a balanced design to achieve the same accuracy for parameter estimation as 100 individuals in a c-optimal design. For heavier censoring, the c-optimal designs are even more preferable.

6.4.2 Analytical results

As mentioned in section 6.1 the partial likelihood function and therefore $\hat{\beta}_{PL}$, depend on the order of occurrence of the event of interest for the various subjects in the experiment. Hence when the partial likelihood method is used we lose information and $\hat{\beta}_{PL}$ is not a sufficient statistic for estimating β. As a result the asymptotic variances of $\hat{\beta}_{PL}$ and $\hat{\beta}_{FL}$ are not equal.

However, the results presented in the previous section suggest that the two variances are close, thus producing similar designs. In what follows, we find an explanation for the similarities of c- and Σ-optimal designs, in particular under heavy censoring and small to moderate β-values.

From section 4.4.3, the Q-function implicitly defined in (4.1) is given by $Q(\alpha+\beta x) = (1 - e^{-ce^{\alpha+\beta x}})$ for the exponential-based proportional hazards model model with Type-I censoring. As shown in Lemma 2, section 4.3.2, the locally c-optimal design for estimating β is always supported at exactly two points. Let x_1, x_2 be the two support points with corresponding weights $1-\omega$ and ω respectively. For heavy censoring, and thus small values of c, the asymptotic variance of ($\sqrt{n}$ times) the maximum full likelihood estimator for β, $\hat{\beta}_{FL}$, can be approximated by a first order Taylor expansion given by

$$\begin{aligned} Var(\hat{\beta}_{FL}) &= \frac{(1-\omega)(1-e^{-ce^{\alpha+\beta x_1}}) + \omega(1-e^{-ce^{\alpha+\beta x_2}})}{\omega(1-\omega)(1-e^{-ce^{\alpha+\beta x_1}})(1-e^{-ce^{\alpha+\beta x_2}})(x_2-x_1)^2} \\ &\approx \frac{(1-\omega)e^{\beta x_1} + \omega e^{\beta x_2}}{\omega(1-\omega)ce^{\alpha}e^{\beta(x_1+x_2)}(x_2-x_1)^2}, \end{aligned}$$

using that

$$1 - e^{-ce^{\alpha+\beta x}} \approx ce^{\alpha+\beta x}.$$

The smaller the value of $ce^{\alpha+\beta x}$, the more accurate the approximation.

Now consider the corresponding quantity for the partial likelihood model for two different treatments or drug doses x_1 and x_2. Without loss of generality we assume that among the data available for n subjects, there are k distinct event times, $t_1 < \ldots < t_k$. Also let r_j be the number of individuals in the risk set at time t_j, q_j of them allocated at x_2 and $r_j - q_j$ allocated at x_1. Then the log-likelihood function defined in (6.3) becomes

$$\log L_P(\beta) = \sum_{j=1}^{k} \left\{ \beta x_j - \log\left[(r_j - q_j)e^{\beta x_1} + q_j e^{\beta x_2}\right]\right\}.$$

Taking the second derivative of the above expression with respect to β, the asymptotic variance of $\sqrt{n}\hat{\beta}_{PL}$, $Var(\hat{\beta}_{PL})$, becomes

$$Var(\hat{\beta}_{PL}) = \left[\frac{1}{n} E \left(\sum_{j=1}^{k} \frac{q_j(r_j - q_j)e^{\beta(x_1+x_2)}(x_2 - x_1)^2}{[(r_j - q_j)e^{\beta x_1} + q_j e^{\beta x_2}]^2} \right) \right]^{-1}. \tag{6.14}$$

Let $q_j^* = q_j/r_j$ and $r_j^* = r_j - q_j/r_j = 1$, $j = 1, \ldots, k$. Then the right hand side of (6.14) will not change when replacing q_j and r_j with q_j^* and r_j^*, respectively. When k/n is small, this means that the proportion of observed event times is small and this corresponds to the case of heavy censoring. Therefore, $q_j^* \approx \omega$ and $r_j^* - q_j^* \approx 1 - \omega$ that is, the original proportion of subjects allocated at x_2 and x_1 respectively, at least for small j. Similarly, if $|\beta|$ is small, the proportion of subjects at risk in the two groups will not change much over time, and again $q_j^* \approx \omega$ in this situation.

Now k, the number of observed events, is itself random, and we replace it with its expectation, approximated by $E(k) \approx n[(1-\omega)ce^{\alpha+\beta x_1} + \omega ce^{\alpha+\beta x_2}]$. Overall, we obtain

$$Var(\hat{\beta}_{PL}) \approx \frac{(1-\omega)e^{\beta x_1} + \omega e^{\beta x_2}}{\omega(1-\omega)ce^{\alpha}e^{\beta(x_1+x_2)}(x_2 - x_1)^2}$$

Hence the two variances, and thus the optimal designs, are approximately equal for c and k/n small, which confirms the numerical results in Tables 6.6 and 6.7.

Under random censoring, $Q(\alpha + \beta x) = 1 + (e^{-ce^{\alpha+\beta x}} - 1)/ce^{\alpha+\beta x} \approx ce^{\alpha+\beta x}/2$ (see section 4.4.3). Following along the same lines as for Type-I censoring, we find that for small values of c and k/n

$$Var(\hat{\beta}_{FL}) \approx Var(\hat{\beta}_{PL}) \approx \frac{2((1-\omega)e^{\beta x_1} + \omega e^{\beta x_2})}{\omega(1-\omega)ce^{\alpha}e^{\beta(x_1+x_2)}(x_2 - x_1)^2}.$$

Therefore, again the two asymptotic variances, and thus the corresponding optimal designs, are approximately equal for heavy censoring.

6.5 Conclusions

The partial likelihood approach has become very popular in survival experiments due to the flexibility provided by the Cox model that leaves the baseline hazard function unspecified. Moreover, under the proportional hazards framework the interpretation of the regression coefficients is straightforward as they represent the effect of the corresponding explanatory variables involved in the model.

However, the construction of optimal designs based on this method depends on non-trivial numerical methods and approximations that practitioners may not find appealing to implement and therefore, the use of some "off the shelf" designs might be more preferable. Such designs are the balanced design allocating equal proportions of subjects at the support points and also the locally optimal designs found in Chapter 4 as they are analytically characterised using easy to check conditions on the parameters.

We have shown that the balanced design performs well under both Type-I and random censoring only in the case of small and moderate absolute values of the covariate effect parameter β ($|\beta| \leq 2.3$) and small percentages of censoring, that is, below 50% (see Tables 6.2-6.5 for detailed values of the efficiencies). This situation is often met in studies where, for example, an alternative treatment under investigation improves the health of patients only by a little compared to the treatment currently in use. Therefore, the β-value, that is the effect of the new treatment, and also the percentage of censoring, corresponding to the number of patients whose health has been improved by the end of the experiment, will both be small.

However, in many industrial reliability studies where, for example, the event of interest is the failure of a particular component this might not be the case as often the component under investigation will be extremely reliable thus resulting in heavy censoring. An example of a situation where the absolute value of the β-parameter will be large, is a clinical trial comparing a placebo to a new treatment. If this treatment is effective and is the first one available for the cure of a particular health issue then its effect will be large. In either of these cases the balanced design will not be sufficiently efficient. Therefore, in such scenarios the use of our analytical characterisations of locally optimal designs based on the full likelihood approach is more appropriate as we have shown that they are efficient also for partial likelihood estimation (see Tables 6.6 and 6.7). Furthermore, under Type-I censoring the locally optimal designs for the exponential regression model can be used without loss of generality as these are optimal independent of the form of the baseline hazard (see discussion in section 6.3.2).

Chapter 7

Model-robust designs

The optimal designs presented in Chapters 4 and 5 were constructed assuming that the model generating the data is known, up to the values of the parameters involved. In many practical situations however, the proposed parametric model will only hold approximately, thus making the formulation of model robust criteria for the construction of optimal designs an issue of great interest and importance. The results discussed in the previous chapter on Cox's proportional hazards model, for which the underlying distribution of the data is left unspecified, indicate that the full and partial likelihood approaches result in very similar designs. Moreover, in the presence of Type-I censoring a constant baseline hazard and therefore the exponential regression model can be assumed without affecting the optimal choice of design.

Based on these observations and taking into account that the exponential distribution is naturally assumed in survival experiments for time to event data, in this chapter we investigate the construction of designs for maximum full likelihood estimation which are robust to misspecifications of the exponential-based proportional hazards model when the data are subject to Type-I censoring. In particular, we consider deviations in a neighbourhood of the exponential regression model that includes widely used parametric proportional hazards models, for example, based on the Weibull and Gompertz distributions. These small deviations which may occur from imprecisions in the specification of the mean response, are formulated directly in the hazard function using what we call a contamination function or simply a contaminant.

In what follows, we first introduce the models considered and define two different classes of contaminants to account for the various possible forms of the functions. We then derive the optimality criteria to be used which correspond to the classical D- and c-optimality criteria but are based on the mean squared error matrix rather than just the information matrix. Finally, minimax designs for maximum likelihood estimation are constructed for binary and continuous design spaces. These designs minimise the corresponding criteria functions for the worst possible contaminant.

7.1 Models and contamination functions

As before, throughout this chapter we assume that the models involve only one explanatory variable taking values x in the design space $\mathcal{X}$ and the aim of the experiment is to estimate one or both of the two model parameters parameters α and β. Also let n be the total number of subjects utilised and c be the duration of the experiment at which point the observations of subjects for which the event of interest has not occurred are said to be right-censored.

We consider the situation where the experimenter assumes the exponential regression model for the censored data which in its proportional hazards parametrisation is specified by the hazard function

$$h_1(t) = \exp\{\alpha + \beta x\}, \quad t > 0, \tag{7.1}$$

when in fact this is only an approximation of the true model which has hazard function given by

$$h_2(t) = \exp\left\{\alpha + \beta x + \frac{g(t)}{\sqrt{n}}\right\}, \quad t > 0. \tag{7.2}$$

The function $g(t)$ represents uncertainty about the exact form of the distribution for the data and we call it the contamination function or just the contaminant. We assume $g(t)$ is unknown and relatively small but we do not estimate it. The factor $n^{-1/2}$ is included so that the deviations are of the order $O(1/\sqrt{n})$ and we have models that are "close to" the exponential regression model.

The parametrisation in (7.2) allows us to remain in a proportional hazards framework in order to ensure that the model parameters are well defined. In particular, the contamination function is independent of the covariate value x and therefore β is the coefficient of the explanatory variable corresponding to the covariate effect. For identifiability reasons we require that $g(t)$ does not involve an additive constant. Otherwise, the constant term would be included in the e^{α} quantity representing the baseline hazard.

We also note that the dependence of g on the time t ensures that the general form of the true model includes widely used parametric proportional hazards models based, for example, on the Weibull and Gompertz distributions with known shape parameter γ for which $g(t)$ is equal to $(\gamma - 1)\ln t$ and γt respectively.

We now define two classes of contamination functions which can be used to include various forms of g. These classes are specified so that the worst possible contaminant can be identified for use in the construction of minimax designs.

The first class of contaminants is specified by

$$\mathcal{G}_1 = \left\{ g : \max_{t\in[0,c]} |g(t)| \leq c_1 \right\}, \tag{7.3}$$

where c_1 is a positive constant assumed to be known. This class includes contamination functions $g(t)$ which are bounded on the time interval $[0, c]$.

Now consider the case of unbounded contamination functions like, for example, $g(t) = \log t$ for which $\lim_{t\to 0} g(t) = -\infty$. A class that can be used to include such contaminants is

$$\mathcal{G}_2 = \left\{ g : \left| \int_0^c e^{-te^{\alpha+\beta x}} g(t)dt \right| \leq c_2, \quad \forall x \in \mathcal{X} \text{ and } \left| \int_0^c g(s)\, ds \right| < \infty \right\}, \tag{7.4}$$

where again c_2 is a known positive constant.

The integral expression involved in the definition of $\mathcal{G}_2$ appears in the optimality criteria functions discussed in the next section. Therefore, if we assume g belongs in the class $\mathcal{G}_2$ this expression has a fixed and known upper bound given by c_2. If $g \in \mathcal{G}_1$ the worst possible value for the contamination function is equal to c_1 and the upper bound of the integral expression depends on the value x of the explanatory variable.

7.2 Optimality criteria

As mentioned previously, here we adopt the full likelihood approach since the assumed parametric model is completely specified as the exponential based proportional hazards model. However, fitting model (7.1) when in fact the true model is given by (7.2) adds a bias to the maximum likelihood estimator vector of the model parameters besides its natural variation.

We therefore consider criteria based on the mean squared error matrix rather than just the information matrix which correspond to the classical criteria of D- and c-optimality defined in sections 4.2.1 and 4.3.1 of Chapter 4. D-optimality for the mean squared error matrix is one of the criteria used by Wiens (1992) who looks at approximately linear regression models for complete data, that is, in the absence of censoring. As he points out, the advantages of using an optimal design minimising just the variance are lost even if the deviations are very small.

In this section we first present some mathematical preliminaries required for the derivation of the mean squared error matrix which is then explicitly determined. Finally we define the criteria functions to be used for the construction of minimax optimal two-point designs.

7.2.1 Preliminaries

Xu (2009b) considers the construction of designs which are robust to misspecifications in the regression function of generalised linear models for censored data with normal underlying distribution. He focuses on prediction and extrapolation problems, thus making the optimality criteria used inapplicable for the estimation problem we consider here. However, throughout this section we follow a similar procedure to the one used in Xu (2009b) for the derivation of expectations and covariances required for constructing the mean squared error matrix.

Let $T_1, \ldots, T_n$ be independent random variables indicating the times to event for the n subjects utilised in the experiment with corresponding observed values $t_1, \ldots, t_n$ and $Y_j = \max\{T_j, c\}$, $j = 1, \ldots, n$ be the random variables for the actual observations in the presence of Type-I censoring. Also let δ_j be an indicator function taking the value 1 when the jth observation is not censored and 0 otherwise. That is

$$\delta_j = \begin{cases} 1, & \text{if } Y_j = T_j \\ 0, & \text{if } Y_j = c \end{cases} .$$

The probability density and survivor functions for the exponential regression model defined in (7.1) are

$$f_1(y_j) = e^{\alpha+\beta x_j} e^{-y_j e^{\alpha+\beta x_j}} \quad \text{and} \quad S_1(y_j) = e^{-ce^{\alpha+\beta x_j}}, \quad j = 1, \ldots, n$$

respectively. Therefore, assuming that model (7.1) is correct, the log-likelihood function of the jth observation with covariate vector x_j is given by

$$\begin{aligned} l(x_j, \alpha, \beta) &= \delta_j \log f_1(y_j) + (1 - \delta_j) \log S_1(c) \\ &= \delta_j \left(\alpha + \beta x_j - y_j e^{\alpha+\beta x_j}\right) - (1 - \delta_j) c e^{\alpha+\beta x_j}. \end{aligned}$$

Taking the first and second order derivatives of this log-likelihood function with respect to the model parameters α and β, we have

$$\frac{\partial l(x_j, \alpha, \beta)}{\partial \alpha} = \delta_j \left(1 - y_j e^{\alpha+\beta x_j}\right) - (1 - \delta_j) c e^{\alpha+\beta x_j}, \qquad \frac{\partial l(x_j, \alpha, \beta)}{\partial \beta} = x_j \frac{\partial l(x_j, \alpha, \beta)}{\partial \alpha}$$

and

$$\frac{\partial^2 l(x_j, \alpha, \beta)}{\partial \alpha^2} = -e^{\alpha+\beta x_j} \left[\delta_j y_j + c(1 - \delta_j)\right], \qquad \frac{\partial^2 l(x_j, \alpha, \beta)}{\partial \alpha \partial \beta} = x_j \frac{\partial^2 l(x_j, \alpha, \beta)}{\partial \alpha^2},$$

$$\frac{\partial^2 l(x_j, \alpha, \beta)}{\partial \beta^2} = x_j^2 \frac{\partial^2 l(x_j, \alpha, \beta)}{\partial \alpha^2}.$$

The above expressions only involve two random quantities via δ_j and $\delta_j Y_j$. We note that $\delta_j \sim Bin(1, P_j)$, where $P_j = P(\delta_i = 1) = P(Y_j = T_j)$ and

$$\delta_j Y_j = \begin{cases} Y_j, & \text{if } Y_j = T_j \\ 0, & \text{if } Y_j = c \end{cases} .$$

We now take into account that the true model is actually specified by (7.2) with corresponding probability density function given by

$$f_2(y_j) = \exp\left\{\alpha + \beta x_j + \frac{g(y_j)}{\sqrt{n}}\right\} \exp\left\{-e^{\alpha+\beta x_j} \int_0^{y_j} e^{g(s)/\sqrt{n}}\, ds\right\}, \quad j = 1, \ldots, n.$$

Based on this true model we derive the expectations and variances of the random variables δ_j, $\delta_j Y_j$ separately and also find their covariance using two Taylor expansions. The calculations are given in detail only for the expectation of δ_j and the rest of the expressions can be proved following similar arguments.

$$\begin{aligned} E[\delta_j] = P_j &= \int_0^c \exp\left\{\alpha + \beta x_j + \frac{g(y_j)}{\sqrt{n}}\right\} \exp\left\{-e^{\alpha+\beta x_j} \int_0^{y_j} e^{g(s)/\sqrt{n}}\, ds\right\} dy_j \\ &= 1 - \exp\left\{-e^{\alpha+\beta x_j} \int_0^c e^{g(s)/\sqrt{n}}\, ds\right\}. \end{aligned}$$

Since we consider small deviations we can take the Taylor expansion of $e^{g(s)/\sqrt{n}}$ around $g(s) = 0$. Then the above expression becomes

$$\begin{aligned} E[\delta_j] &= 1 - \exp\left\{-e^{\alpha+\beta x_j} \int_0^c 1 + \frac{g(s)}{\sqrt{n}} + o\left(\frac{g(s)}{\sqrt{n}}\right) ds\right\} \\ &= 1 - \exp\left\{-ce^{\alpha+\beta x_j}\right\} \exp\left\{-e^{\alpha+\beta x_j}\left[\int_0^c \frac{g(s)}{\sqrt{n}}\, ds + o\left(\frac{1}{\sqrt{n}}\right)\right]\right\}. \end{aligned}$$

By further expanding around $\int_0^c \frac{g(s)}{\sqrt{n}}\, ds + o\left(\frac{1}{\sqrt{n}}\right) = 0$, we find that the expectation of the random variable δ_j is

$$E[\delta_j] = 1 - e^{-ce^{\alpha+\beta x_j}} + e^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}} \int_0^c \frac{g(s)}{\sqrt{n}}\, ds + o\left(\frac{1}{\sqrt{n}}\right),$$

where the first term, $1 - e^{-ce^{\alpha+\beta x_i}}$, corresponds to the expectation if the exponential regression model was in fact the true model. Using the above expression the variance of δ_j can be found without making any further calculations and is given by

$$\begin{aligned} Var(\delta_j) = P_j(1-P_j) &= e^{-ce^{\alpha+\beta x_j}}(1 - e^{-ce^{\alpha+\beta x_j}}) \\ &\quad + e^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}}(2e^{-ce^{\alpha+\beta x_j}} - 1) \int_0^c \frac{g(s)}{\sqrt{n}}\, ds + o\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Following along the same lines as for $E[\delta_j]$, we obtain

$$\begin{aligned} E[\delta_j Y_j] = &\frac{(1 - e^{-ce^{\alpha+\beta x_j}})}{e^{\alpha+\beta x_j}} - ce^{-ce^{\alpha+\beta x_j}} + e^{-ce^{\alpha+\beta x_j}}(ce^{\alpha+\beta x_j} + 1) \int_0^c \frac{g(s)}{\sqrt{n}}\, ds \\ &- \int_0^c \frac{g(y_i)}{\sqrt{n}} e^{-y_j e^{\alpha+\beta x_j}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

$$
\begin{aligned}
Var(\delta_j Y_j) =& - c^2 e^{-ce^{\alpha+\beta x_j}}(1+e^{-ce^{\alpha+\beta x_j}}) + \frac{(1-e^{-2ce^{\alpha+\beta x_j}})}{(e^{\alpha+\beta x_j})^2} - \frac{2ce^{-2ce^{\alpha+\beta x_j}}}{e^{\alpha+\beta x_j}} + e^{-ce^{\alpha+\beta x_j}} \\
& \left(c^2 e^{\alpha+\beta x_j} + 4ce^{-ce^{\alpha+\beta x_j}} + \frac{2e^{-ce^{\alpha+\beta x_j}}}{e^{\alpha+\beta x_j}} + 2c^2 e^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}}\right) \int_0^c \frac{g(s)}{\sqrt{n}}\, ds \\
& - \int_0^c 2e^{-y_j e^{\alpha+\beta x_j}} \left(y_j + \frac{e^{-ce^{\alpha+\beta x_j}}}{e^{\alpha+\beta x_j}} + ce^{-ce^{\alpha+\beta x_j}}\right) \frac{g(y_j)}{\sqrt{n}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

$$
\begin{aligned}
Cov(\delta_j, \delta_j Y_j) =& \, e^{-ce^{\alpha+\beta x_j}}(1-e^{-ce^{\alpha+\beta x_j}})/e^{\alpha+\beta x_j} - ce^{-2ce^{\alpha+\beta x_j}} \\
& + e^{-ce^{\alpha+\beta x_j}} \left(2ce^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}} + 2e^{-ce^{\alpha+\beta x_j}} - 1\right) \int_0^c \frac{g(s)}{\sqrt{n}}\, ds \\
& - e^{-ce^{\alpha+\beta x_j}} \int_0^c \frac{g(y_j)}{\sqrt{n}} e^{-y_j e^{\alpha+\beta x_j}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Hence

$$
E\left[\frac{\partial l(x_j,\alpha,\beta)}{\partial \alpha}\right] = e^{\alpha+\beta x_j} \int_0^c e^{-y_j e^{\alpha+\beta x_j}} \frac{g(y_j)}{\sqrt{n}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right),
$$

$$
\begin{aligned}
E\left[-\frac{\partial^2 l(x_j,\alpha,\beta)}{\partial \alpha^2}\right] =& \, 1 - e^{-ce^{\alpha+\beta x_j}} + e^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}} \int_0^c \frac{g(s)}{\sqrt{n}}\, ds \\
& - e^{\alpha+\beta x_j} \int_0^c \frac{g(y_j)}{\sqrt{n}} e^{-y_j e^{\alpha+\beta x_j}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

$$
\begin{aligned}
Var\left(\frac{\partial l(x_j,\alpha,\beta)}{\partial \alpha}\right) =& \, 1 - e^{-ce^{\alpha+\beta x_j}} + e^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}} \int_0^c \frac{g(s)}{\sqrt{n}}\, ds \\
& - (e^{\alpha+\beta x_j})^2 \int_0^c 2\frac{y_j g(y_j)}{\sqrt{n}} e^{-y_j e^{\alpha+\beta x_j}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

$$
\begin{aligned}
Cov\left(\frac{\partial l(x_j,\alpha,\beta)}{\partial \alpha}, \frac{\partial l(x_j,\alpha,\beta)}{\partial \beta}\right) &= x_j Var\left(\frac{\partial l(x_j,\alpha,\beta)}{\partial \alpha}\right), \\
Var\left(\frac{\partial l(x_j,\alpha,\beta)}{\partial \beta}\right) &= x_j^2 Var\left(\frac{\partial l(x_j,\alpha,\beta)}{\partial \alpha}\right).
\end{aligned}
$$

7.2.2 Mean squared error matrix

Let $\boldsymbol{\lambda} = (\alpha, \beta)^T$ be the vector of the model parameters and $\boldsymbol{\lambda_0}$ the vector of their true values. Also let ξ be an approximate design supported at points $x_1, \ldots, x_m$ taking values in the design space $\mathcal{X}$ with corresponding weights $\omega_1, \ldots, \omega_m$ where $0 < \omega_i \leq 1$, $i = 1, \ldots, m$ and $\sum_{i=1}^m \omega_i = 1$. Using the results presented in the previous section we obtain the following expressions.

The asymptotic information matrix of $\boldsymbol{\lambda_0}$ is

$$\begin{aligned} M(\xi) = M(\xi, \boldsymbol{\lambda_0}) &= \lim_{n\to\infty} \frac{1}{n} E\left[-\sum_{j=1}^{n} \frac{\partial^2 l(x_j,\alpha,\beta)}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda^T}}\Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda_0}} \right] \\ &= \sum_{i=1}^{m} \omega_i (1 - e^{-ce^{\alpha+\beta x_i}}) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}, \end{aligned}$$

the asymptotic expectation of the score function evaluated at $\boldsymbol{\lambda_0}$ is

$$\begin{aligned} \tilde{b}(\xi, g) = \tilde{b}(\xi, g, \boldsymbol{\lambda_0}) &= \frac{1}{\sqrt{n}} \lim_{n\to\infty} \frac{1}{n} E\left[\sqrt{n} \sum_{j=1}^{n} \frac{\partial l(x_j,\alpha,\beta)}{\partial\boldsymbol{\lambda}}\Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda_0}} \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{m} \omega_i e^{\alpha+\beta x_i} \int_0^c e^{-y_j e^{\alpha+\beta x_i}} g(y_j)\, dy_j \begin{pmatrix} 1 \\ x_i \end{pmatrix} := \frac{1}{\sqrt{n}} b(\xi, g), \end{aligned}$$

and finally the asymptotic variance-covariance matrix of the score function $S(\boldsymbol{\lambda})$ evaluated at $\boldsymbol{\lambda_0}$ is given by

$$\begin{aligned} C(\xi) = C(\xi, \boldsymbol{\lambda_0}) &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} Cov\left(\frac{\partial l(x_j,\alpha,\beta)}{\partial\boldsymbol{\lambda}}\Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda_0}} \right) \\ &= \sum_{i=1}^{m} \omega_i (1 - e^{-ce^{\alpha+\beta x_i}}) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}. \end{aligned}$$

In order to obtain the mean squared error matrix we first have to find the asymptotic distribution of $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda_0}$. Expanding the score function around $\boldsymbol{\lambda_0}$ gives

$$S(\boldsymbol{\lambda}) = S(\boldsymbol{\lambda_0}) + S'(\boldsymbol{\lambda_0})(\boldsymbol{\lambda} - \boldsymbol{\lambda_0}) + \ldots,$$

and using the fact that the maximum likelihood estimate $\hat{\boldsymbol{\lambda}}$ is a root of the score function we have

$$\begin{aligned} 0 &\approx S(\boldsymbol{\lambda_0}) + S'(\boldsymbol{\lambda_0})(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda_0}) \\ (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda_0}) &\approx M^{-1}(\xi, \boldsymbol{\lambda_0}) S(\boldsymbol{\lambda_0}). \end{aligned}$$

Now $\sqrt{n}\, S(\boldsymbol{\lambda_0}) \sim AN(b, C)$ and therefore

$$\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda_0}) \sim AN(M^{-1}b, M^{-1}CM^{-1}).$$

Hence the mean squared error matrix of $\sqrt{n}\,(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda_0})$ is given by

$$\begin{aligned} MSE(\xi, \alpha, \beta, g) &= (M^{-1}b)(M^{-1}b)^T + M^{-1}CM^{-1} \\ &= M^{-1}(bb^T + C)M^{-1}, \end{aligned} \tag{7.5}$$

since the asymptotic information matrix M is symmetric.

7.2.3 Minimax designs

We now present the optimality criteria to be used for the construction of minimax designs. These designs minimise the criteria functions of the mean squared error matrix corresponding to the classical D- and c-optimality criteria, for the worst possible contamination function g. Therefore, we first fix a design ξ and maximise the criteria functions over $\mathcal{G}_1$ or $\mathcal{G}_2$ and finally minimise the resulting expression with respect to the design. In what is to follow we consider approximate designs with two support points and so the criteria are given for two-point designs.

The first criterion we consider corresponds to the c-optimality criterion for estimating the parameter β. The estimation of this parameter is of primary interest in survival experiments since it represents the covariate effect. We call ξ^* a minimax c-optimal design for estimating β if $(0\ \ 1)^T$ is in the range of $MSE(\xi^*, \alpha, \beta, g)$ and

$$\xi^* = \arg\min_{\xi} \max_{g \in \mathcal{G}_1 or \mathcal{G}_2} (0 \quad 1)\, MSE^{-1}(\xi, \alpha, \beta, g) \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{7.6}$$

Similarly a design ξ^* is minimax c-optimal for estimating α if $(0\ \ 1)^T$ is in the range of $MSE(\xi^*, \alpha, \beta, g)$ and

$$\xi^* = \arg\min_{\xi} \max_{g \in \mathcal{G}_1 or \mathcal{G}_2} (1 \quad 0)\, MSE^{-1}(\xi, \alpha, \beta, g) \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{7.7}$$

As discussed in Chapter 5, the parameter α is involved in the baseline hazard function and therefore in practice a reasonable approximation of its value might be available. The reason for considering c-optimality for α is given in section 7.3.2.

Finally, we find minimax D-optimal designs ξ^* where

$$\xi^* = \arg\min_{\xi} \max_{g \in \mathcal{G}_1 or \mathcal{G}_2} det \left\{ MSE(\xi, \alpha, \beta, g) \right\}. \tag{7.8}$$

This criterion is used when we are interested in estimating both of the model parameters α and β. We note that, unlike the classical D-optimality criterion where ξ^* maximises the determinant of the information matrix with respect to the design ξ, here we find optimal designs that minimise the maximum value of the determinant of the mean squared error matrix over all probability measures ξ.

7.3 Minimax optimal two-point designs for binary design space

When the experiment is conducted to compare, for example, two treatments the design space is binary and given by $\mathcal{X} = \{0, 1\}$. In this case, the designs are supported at points 0 and 1 with corresponding weights ω and $1-\omega$ where $0 < \omega \leq 1$ and the expressions found in section 7.2.2 for the asymptotic information and variance-covariance matrices and for the ($\sqrt{n}$ times) asymptotic expectation of the score function become

$$M(\xi) = C(\xi) = \begin{pmatrix} \omega(1-e^{ce^{\alpha}}) + (1-\omega)(1-e^{-ce^{\alpha+\beta}}) & (1-\omega)(1-e^{-ce^{\alpha+\beta}}) \\ (1-\omega)(1-e^{-ce^{\alpha+\beta}}) & (1-\omega)(1-e^{-ce^{\alpha+\beta}}) \end{pmatrix}$$

$$b(\xi, g) = \begin{pmatrix} \omega e^{\alpha} \int_0^c e^{-y_j e^{\alpha}} g(y_j)\, dy_j + (1-\omega) e^{\alpha+\beta} \int_0^c e^{-y_j e^{\alpha+\beta}} g(y_j)\, dy_j \\ (1-\omega) e^{\alpha+\beta} \int_0^c e^{-y_j e^{\alpha+\beta}} g(y_j)\, dy_j \end{pmatrix}.$$

The mean squared error matrix is then given by $M^{-1}(bb^T + C)M^{-1}$ as shown in section 7.2.2. The full matrix is not illustrated here due to its large size.

In what follows we present the functions to be optimised for each of the optimality criteria using the definitions given in (7.6), (7.7) and (7.8) and the resulting minimax optimal designs.

7.3.1 Minimax c-optimal designs for β

The objective function defined in (7.6) is given by

$$\left[\frac{e^{\alpha+\beta}}{(1-e^{-ce^{\alpha+\beta}})} \int_0^c e^{-y_j e^{\alpha+\beta}} g(y_j)\, dy_j - \frac{e^{\alpha}}{(1-e^{-ce^{\alpha}})} \int_0^c e^{-y_j e^{\alpha}} g(y_j)\, dy_j \right]^2 + \frac{1}{\omega(1-e^{-ce^{\alpha}})} + \frac{1}{\omega(1-e^{-ce^{\alpha+\beta}})}.$$

The minimax c-optimal design for β is found by minimising the above expression with respect to ω for the worst possible contaminant.

It is easy to see that the term involving the contamination function g is independent of the weight ω. Therefore, it is enough to minimise

$$\frac{1}{\omega(1-e^{-ce^{\alpha}})} + \frac{1}{\omega(1-e^{-ce^{\alpha+\beta}})},$$

which gives the optimal weight

$$\omega^* = \frac{\sqrt{1-e^{-ce^{\alpha+\beta}}}}{\sqrt{1-e^{-ce^{\alpha}}} + \sqrt{1-e^{-ce^{\alpha+\beta}}}}. \tag{7.9}$$

This is the same c-optimal weight as for the case of the exponential-based proportional hazards model being the true model. Therefore, the contamination function does not affect the minimax c-optimal design for estimating β and the exponential regression model can be assumed without loss of generality. Table 4.3 in section 4.4.4 of Chapter 4 presents these weights for various parameter values.

This result comes to an agreement with the conclusions found in the previous chapter for partial likelihood estimation. That is, in the presence of Type-I censoring the optimal choice of design is independent of the parametric model used, if the proportional hazards assumption is true.

7.3.2 Minimax c-optimal designs for α

Based on the results found above we felt it is worthwhile examining whether the contamination function plays a role in the construction of the minimax optimal design when interest centres on estimating only the parameter α, even though in practice this will rarely be the case.

We therefore also consider the construction of minimax c-optimal designs for α which minimise the function

$$\left[\frac{e^{\alpha}}{(1-e^{-ce^{\alpha}})}\int_0^c e^{-y_j e^{\alpha}} g(y_j)\,dy_j\right]^2 + \frac{1}{\omega(1-e^{-ce^{\alpha}})},$$

with respect to the weight ω for the worst possible g.

As before, the contamination function does not affect the optimal design since we only have to minimise the expression $1/\omega(1-e^{-ce^{\alpha}})$ which gives $\omega^* = 1$. Hence the minimax c-optimal design for α is a one-point design putting all observations at point 0. This is on accordance with the optimal design for the linear model when we are interested in estimating only the intercept.

7.3.3 Minimax D-optimal designs

For a fixed design in $\mathcal{X} = \{0, 1\}$ supported at 0 and 1 with corresponding weights ω and $1-\omega$ the determinant of the mean squared error matrix is given by

$$\frac{1}{\omega(1-\omega)(1-e^{-ce^{\alpha}})(1-e^{-ce^{\alpha+\beta}})}\left\{1+\omega\frac{\left[e^{\alpha}\int_0^c e^{-y_j e^{\alpha}} g(y_j)\,dy_j\right]^2}{(1-e^{-ce^{\alpha}})}\right.$$
$$\left. +\,(1-\omega)\frac{\left[e^{\alpha+\beta}\int_0^c e^{-y_j e^{\alpha+\beta}} g(y_j)\,dy_j\right]^2}{(1-e^{-ce^{\alpha+\beta}})}\right\}. \qquad (7.10)$$

Minimax D-optimal designs are constructed by minimising, with respect to ω, the maximum of the above expression taken over the class of possible contamination functions g. We now consider the two different classes of contaminants separately defined in (7.3) and (7.4).

The following theorem gives the minimax D-optimal weight corresponding to point $x = 0$ assuming that the contamination function belongs in the class $\mathcal{G}_1$.

Theorem 7. *Let $g \in \mathcal{G}_1$ and $\mathcal{X} = \{0,1\}$. The minimax D-optimal two-point design supported at points 0 and 1 allocates a proportion ω^* of observations at point 0 regardless of the sign of the parameter β, where*

$$\omega^* = \frac{\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta}})+1}\left[\sqrt{c_1^2(1-e^{-ce^{\alpha}})+1}-\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta}})+1}\right]}{c_1^2(e^{-ce^{\alpha+\beta}}-e^{-ce^{\alpha}})}. \tag{7.11}$$

Proof. If $g \in \mathcal{G}_1$ then $\max_{y_j\in[0,c]} |g(y_j)| \le c_1 \ \forall j = 1,\ldots,n$ and so

$$\left|\int_0^c e^{-y_j e^{\alpha+\beta x}} g(y_j)\, dy_j\right| \le \int_0^c e^{-y_j e^{\alpha+\beta x}} |g(y_j)|\, dy_j \le \int_0^c e^{-y_j e^{\alpha+\beta x}} c_1\, dy_j$$
$$= c_1(1-e^{-ce^{\alpha+\beta x}})/e^{\alpha+\beta x}, \qquad \forall x \in \{0,1\}$$

Therefore, for contamination functions g in the class $\mathcal{G}_1$ the maximum value of (7.10) is given by

$$\frac{c_1^2}{\omega(1-e^{-ce^{\alpha}})}+\frac{c_1^2}{(1-\omega)(1-e^{-ce^{\alpha+\beta}})}+\frac{1}{\omega(1-\omega)(1-e^{-ce^{\alpha}})(1-e^{-ce^{\alpha+\beta}})}.$$

Taking the first order derivative of this expression with respect to ω and equating it to zero gives

$$c_1^2\omega^2(1-e^{-ce^{\alpha}})-c_1^2(1-\omega)^2(1-e^{-ce^{\alpha+\beta}})-(1-2\omega)=0$$
$$\iff \omega_{1,2} = \frac{-[c_1^2(1-e^{-ce^{\alpha+\beta}})+1] \pm \sqrt{c_1^2(1-e^{-ce^{\alpha}})+1}\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta}})+1}}{c_1^2(e^{-ce^{\alpha+\beta}}-e^{-ce^{\alpha}})}.$$

When β is positive, it is easy to see that both the numerator and the denominator of the above expression are non-positive. We reject the negative root of the numerator as

$$-c_1^2(1-e^{-ce^{\alpha+\beta}})-1-\sqrt{c_1^2(1-e^{-ce^{\alpha}})+1}\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta}})+1}$$
$$< -c_1^2(1-e^{-ce^{\alpha+\beta}})+c_1^2(1-e^{-ce^{\alpha}}) = c_1^2(e^{-ce^{\alpha+\beta}}-e^{-ce^{\alpha}})$$

and the weight must always be always less than or equal to unity.

In the case of negative β-values the denominator is positive and since $\omega > 0$, again we accept the positive root.

Therefore for any sign of the parameter β the minimax D-optimal weight at point 0 is always given by (7.11).

□

The corresponding result for contamination functions $g \in \mathcal{G}_2$ is presented in Theorem 8. This is proven in section B.3.1 of Appendix B following along the same lines.

Theorem 8. *Let $g \in \mathcal{G}_2$ and $\mathcal{X} = \{0, 1\}$. The minimax D-optimal two-point design supported at points 0 and 1 allocates a proportion ω^* of observations at point 0 regardless of the sign of the parameter β, where*

$$\omega^* = \frac{\sqrt{\frac{c_2^2(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}+1}\left[\sqrt{\frac{c_2^2(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}+1}-\sqrt{\frac{c_2^2(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}+1}\right]}{c_2^2\left[\frac{(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}-\frac{(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right]}. \tag{7.12}$$

7.4 Minimax optimal two-point designs for continuous design space

We now consider the case of an arbitrary continuous design space $\mathcal{X} = [u, v]$ corresponding to explanatory variable values indicating, for example, drug doses. We fix a two-point design ξ with x_1, x_2 as its support points and $\omega, 1-\omega$ the corresponding weights and construct the minimax designs in a systematic way by minimising the criteria functions given in (7.6), (7.7) and (7.8) assuming g belongs in the class $\mathcal{G}_1$ or $\mathcal{G}_2$ defined in (7.3) and (7.4) respectively.

Applying the two-point design ξ to the expressions found for the asymptotic matrices $M(\xi)$, $b(\xi, g)$ and $C(\xi)$ we obtain the mean squared error matrix defined in (7.5) which as before is not presented here due to its large size.

7.4.1 Minimax c-optimal designs for β

For the construction of minimax c-optimal design for β we must minimise the function

$$\frac{1}{(x_1-x_2)^2}\left\{\left[\frac{e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x1}}g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}-\frac{e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x2}}g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}\right]^2 + \frac{1}{\omega(1-e^{-ce^{\alpha+\beta x_1}})}+\frac{1}{(1-\omega)(1-e^{-ce^{\alpha+\beta x_2}})}\right\},$$

for the worst contaminant g with respect to x_1, x_2 and ω. We observe that for fixed x_1 and x_2 the minimisation with respect to ω is independent of g and therefore the minimax c-optimal weight ω^* corresponding to the smaller support point is of the same form as for the binary case and is given by

$$\omega^* = \frac{\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}}{\sqrt{1-e^{-ce^{\alpha+\beta x_1}}}+\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}} := \frac{\sqrt{Q(\alpha+\beta x_2)}}{\sqrt{Q(\alpha+\beta x_1)}+\sqrt{Q(\alpha+\beta x_2)}}, \tag{7.13}$$

where the Q-function is the one defined in Chapter 4. However, the contamination function affects the choice of optimal support points and we therefore consider the two classes $\mathcal{G}_1$ and $\mathcal{G}_2$ separately.

An analytical characterisation of minimax c-optimal two-point designs for β when $g \in \mathcal{G}_1$ is provided in Theorem 9 . A sketch proof of part (a) is given below and part (b) is proven in section B.3.2 of Appendix B following along the same lines.

Theorem 9. *Let $g \in \mathcal{G}_1$*
(a) If $\beta > 0$, the design with support points x_1^* and v and optimal weight on x_1^* given in (7.13) is minimax c-optimal for β on $\mathcal{X} = [u, v]$, where $x_1^* = u$ if

$$\beta(u-v) + \frac{2Q(\alpha+\beta u)}{Q'(\alpha+\beta u)}\left[1+\frac{\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta v)}}+\frac{4c_1^2 Q(\alpha+\beta u)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta v)}}\right] > 0. \tag{7.14}$$

Otherwise x_1^* is the unique solution of the equation

$$\beta(x_1-v) + \frac{2Q(\alpha+\beta x_1)}{Q'(\alpha+\beta x_1)}\left[1+\frac{\sqrt{Q(\alpha+\beta x_1)}}{\sqrt{Q(\alpha+\beta v)}}+\frac{4c_1^2 Q(\alpha+\beta x_1)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta x_1)}+\sqrt{Q(\alpha+\beta v)}}\right] = 0. \tag{7.15}$$

(b) If $\beta < 0$, the design with support points u and x_2^* and optimal weight on u given in (7.13) is minimax c-optimal for β on $\mathcal{X} = [u, v]$, where $x_2^* = v$ if

$$\beta(u-v) - \frac{2Q(\alpha+\beta u)}{Q'(\alpha+\beta u)}\left[1+\frac{\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta v)}}+\frac{4c_1^2 Q(\alpha+\beta u)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta v)}}\right] < 0.$$

Otherwise x_2^* is the unique solution of the equation

$$\beta(u-x_2) - \frac{2Q(\alpha+\beta x_2)}{Q'(\alpha+\beta x_2)}\left[1+\frac{\sqrt{Q(\alpha+\beta x_2)}}{\sqrt{Q(\alpha+\beta u)}}+\frac{4c_1^2 Q(\alpha+\beta x_2)\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta x_2)}}\right] = 0.$$

Proof. Let $\beta > 0$. Since $g \in \mathcal{G}_1$ then

$$\left|\int_0^c e^{-y_j e^{\alpha+\beta x}} g(y_j)\, dy_j\right| \le \int_0^c e^{-y_j e^{\alpha+\beta x}} |g(y_j)|\, dy_j \le \int_0^c e^{-y_j e^{\alpha+\beta x}} c_1\, dy_j$$
$$= c_1(1-e^{-ce^{\alpha+\beta x}})/e^{\alpha+\beta x}, \quad \forall x \in [u, v]$$

Therefore,

$$\left[\left|\frac{e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x1}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})} - \frac{e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x2}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}\right|\right]^2$$

$$\le \left[\left|\frac{e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x1}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}\right| + \left|\frac{e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x2}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}\right|\right]^2 \le 4c_1^2$$

Substituting the expression for the c-optimal weights from (7.13) we obtain the objective function to be minimised to be

$$\frac{1}{(x_1-x_2)^2}\left\{4c_1^2+\left[\frac{1}{\sqrt{1-e^{-ce^{\alpha+\beta x_1}}}}+\frac{1}{\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}}\right]^2\right\}:=k(x_1,x_2)$$

For fixed $x_1(<x_2)$, $k(x_1,x_2)$ is decreasing with x_2 as the product of two non-negative decreasing functions and therefore attains its minimum at $x_2^*=v$. Now $k(x_1,v)$ has exactly one turning point on $(-\infty,v)$ which is a minimum since

$$\lim_{x_1\to-\infty}k(x_1,v)=\lim_{x_1\to v}k(x_1,v)=\infty.$$

If this minimum is attained outside $[u,v)$ then the smaller support point of the design is u. This occurs if and only if

$$\left.\frac{\partial k(x_1,v)}{\partial x_1}\right|_{x_1=u}>0,$$

which is equivalent to condition (7.14). Otherwise, x_1^* is the smaller support point and can be found by solving

$$\frac{\partial k(x_1,v)}{\partial x_1}=0$$

which is equivalent to solving equation (7.15).

□

Using Theorem 9 the design problem has either been reduced from a three-dimensional to a one-dimensional optimisation problem or has been solved completely, thus reducing the design search substantially.

For contamination functions $g\in\mathcal{G}_2$ a similar analytical result could not be proven. However, using the fact that

$$\left[\left|\frac{e^{\alpha+\beta x_1}\int_0^c e^{-y_je^{\alpha+\beta x1}}g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}-\frac{e^{\alpha+\beta x_2}\int_0^c e^{-y_je^{\alpha+\beta x2}}g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}\right|\right]^2$$

$$\le\left[\left|\frac{e^{\alpha+\beta x_1}\int_0^c e^{-y_je^{\alpha+\beta x1}}g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}\right|+\left|\frac{e^{\alpha+\beta x_2}\int_0^c e^{-y_je^{\alpha+\beta x2}}g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}\right|\right]^2$$
$$\le c_2^2\left[\frac{e^{\alpha+\beta x_1}}{(1-e^{-ce^{\alpha+\beta x_1}})}+\frac{e^{\alpha+\beta x_2}}{(1-e^{-ce^{\alpha+\beta x_2}})}\right]^2,$$

the support points of the minimax c-optimal two-point design for β can be found by minimising the function

$$\frac{1}{(x_1-x_2)^2}\left\{c_2^2\left[\frac{e^{\alpha+\beta x_1}}{(1-e^{-ce^{\alpha+\beta x_1}})}+\frac{e^{\alpha+\beta x_2}}{(1-e^{-ce^{\alpha+\beta x_2}})}\right]^2 + \left[\frac{1}{\sqrt{1-e^{-ce^{\alpha+\beta x_1}}}}+\frac{1}{\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}}\right]^2\right\}. \quad (7.16)$$

7.4.2 Minimax c-optimal designs for α

Fixing a two-point design ξ supported at points x_1, x_2 with corresponding weights $\omega, 1-\omega$, the objective function defined in (7.7) becomes

$$\frac{1}{(x_1-x_2)^2}\left\{\left[\frac{x_1 e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x_2}} g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}-\frac{x_2 e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x_1}} g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}\right]^2 + \frac{x_2^2}{\omega(1-e^{-ce^{\alpha+\beta x_1}})}+\frac{x_1^2}{(1-\omega)(1-e^{-ce^{\alpha+\beta x_2}})}\right\}.$$

The c-optimal weight ω can be found independently of g by minimising the above expression with respect to ω for fixed x_1, x_2 and is given by

$$\omega^* = \frac{x_2\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}}{x_1\sqrt{1-e^{-ce^{\alpha+\beta x_1}}}+x_2\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}} := \frac{x_2\sqrt{Q(\alpha+\beta x_2)}}{x_1\sqrt{Q(\alpha+\beta x_1)}+x_2\sqrt{Q(\alpha+\beta x_2)}}, \quad (7.17)$$

The following theorem provides a complete classification of minimax c-optimal two-point design for α and $g \in \mathcal{G}_1$. Its proof follows similar arguments as for Theorem 9 and is given in section B.3.3 of Appendix B.

Theorem 10. *Let $g \in \mathcal{G}_1$*
(a) If $\beta > 0$, the design with support points x_1^* and v and optimal weight on x_1^* given in (7.17) is minimax c-optimal for α on $\mathcal{X} = [u, v]$, where $x_1^* = u$ if

$$\beta(u-v)+\frac{2Q(\alpha+\beta u)}{Q'(\alpha+\beta u)}\left[1+\frac{\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta v)}}+\frac{2c_1^2(u+v)Q(\alpha+\beta u)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta v)}}\right] > 0.$$

Otherwise x_1^* is the unique solution of the equation

$$\beta(x_1-v)+\frac{2Q(\alpha+\beta x_1)}{Q'(\alpha+\beta x_1)}\left[1+\frac{\sqrt{Q(\alpha+\beta x_1)}}{\sqrt{Q(\alpha+\beta v)}}+\frac{2c_1^2(x_1+v)Q(\alpha+\beta x_1)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta x_1)}+\sqrt{Q(\alpha+\beta v)}}\right] = 0.$$

(b) If $\beta < 0$, the design with support points u and x_2^* and optimal weight on u given in (7.17) is minimax c-optimal for α on $\mathcal{X} = [u, v]$, where $x_2^* = v$ if

$$\beta(u-v)-\frac{2Q(\alpha+\beta u)}{Q'(\alpha+\beta u)}\left[1+\frac{\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta v)}}+\frac{2c_1^2(u+v)Q(\alpha+\beta u)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta v)}}\right]<0.$$

Otherwise x_2^* *is the unique solution of the equation*

$$\beta(u-x_2)-\frac{2Q(\alpha+\beta x_2)}{Q'(\alpha+\beta x_2)}\left[1+\frac{\sqrt{Q(\alpha+\beta x_2)}}{\sqrt{Q(\alpha+\beta u)}}+\frac{2c_1^2(u+x_2)Q(\alpha+\beta x_2)\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta x_2)}}\right]=0.$$

Theorem 10 provides a complete classification of minimax c-optimal two-point designs for estimating α. Based on some easily verifiable conditions on the parameters, the designs are found by optimising just a one variable function. We also note that in the case of the continuous design space being the interval $[0,1]$ and for $\beta<0$, the smaller support point of the design is $x_1^*=0$ and therefore $\omega^*=1$. Hence the minimax c-optimal design for α coincides with the corresponding c-optimal design for a binary design space and is a one-point design that allocates all the observations at point zero.

When $g\in\mathcal{G}_2$ an analytical characterisation of the minimax designs is not available and the support points are found by minimising the function

$$\frac{1}{(x_1-x_2)^2}\left\{c_2^2\left[\frac{x_2e^{\alpha+\beta x_1}}{(1-e^{-ce^{\alpha+\beta x_1}})}+\frac{x_1e^{\alpha+\beta x_2}}{(1-e^{-ce^{\alpha+\beta x_2}})}\right]^2+\left[\frac{x_2}{\sqrt{1-e^{-ce^{\alpha+\beta x_1}}}}+\frac{x_1}{\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}}\right]^2\right\}. \tag{7.18}$$

7.4.3 Minimax D-optimal designs

In the case of an arbitrary two-point design on a continuous design space, the determinant of the mean squared error matrix is given by

$$\frac{1}{\omega(1-\omega)(1-e^{-ce^{\alpha+\beta x_1}})(1-e^{-ce^{\alpha+\beta x_2}})(x_1-x_2)^2}\left\{1+\frac{(1-\omega)}{(1-e^{-ce^{\alpha+\beta x_2}})}\left(e^{\alpha+\beta x_2}\int_0^c e^{-y_je^{\alpha+\beta x_2}}g(y_j)\,dy_j\right)^2+\frac{\omega}{(1-e^{-ce^{\alpha+\beta x_1}})}\left(e^{\alpha+\beta x_1}\int_0^c e^{-y_je^{\alpha+\beta x_1}}g(y_j)\,dy_j\right)^2\right\}.$$

As for the case of a binary design space the optimal weights do not depend on the form of the contamination function g and are given by

$$\omega^*=\frac{\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta x_2}})+1}\left[\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta x_1}})+1}-\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta x_2}})+1}\right]}{c_1^2(e^{-ce^{\alpha+\beta x_2}}-e^{-ce^{\alpha+\beta x_1}})}, \tag{7.19}$$

when the contamination function belongs in $\mathcal{G}_1$ and if $g\in\mathcal{G}_2$

$$\omega^* = \frac{\sqrt{\frac{c_2^2(e^{\alpha+\beta x_2})^2}{(1-e^{-ce^{\alpha+\beta x_2}})}+1}\left[\sqrt{\frac{c_2^2(e^{\alpha+\beta x_1})^2}{(1-e^{-ce^{\alpha+\beta x_1}})}+1}-\sqrt{\frac{c_2^2(e^{\alpha+\beta x_2})^2}{(1-e^{-ce^{\alpha+\beta x_2}})}+1}\right]}{c_2^2\left[\frac{(e^{\alpha+\beta x_1})^2}{(1-e^{-ce^{\alpha+\beta x_1}})}-\frac{(e^{\alpha+\beta x_2})^2}{(1-e^{-ce^{\alpha+\beta x_2}})}\right]}. \qquad (7.20)$$

Therefore the design problem is again reduced to identifying only the support points of the minimax two-point design which can be found numerically by substituting these in the expression for the determinant of the MSE matrix given above and then minimising the resulting quantity for the worst possible contaminant.

7.5 Application to Freireich data

To better illustrate our results we now apply them to the Freireich data Freireich et al. (1963) obtained from a study comparing a placebo with an active treatment for leukemia with the assumed model being the exponential-based proportional hazards model (see section 5.3 in Chapter 5 for more details). For these data the maximum likelihood estimates of the parameters are $\hat{\alpha} = -2.163$ and $\hat{\beta} = -1.526$ and approximately 30% of the observations are right-censored.

We first consider a binary design space $\mathcal{X} = \{0, 1\}$. The "amount of censoring" for the exponential regression model defined in Kalish and Harrington (1988) as the overall probability of censoring if a balanced design is used, is given by

$$1 - 0.5(1 - e^{-ce^{\alpha+\beta}}) - 0.5(1 - e^{-ce^{\alpha}}).$$

Using the percentage of censoring and the parameter estimates from the Freireich data, this yields $c = 30$.

As mentioned in section 7.3.2, the minimax c-optimal design for α allocates all of the subjects to the experimental point 0 regardless of the parameter values. For $\alpha = -2.163$, $\beta = -1.526$ and $c = 30$ the minimax c-optimal weight for estimating β, given in (7.9), is equal to 0.42. This can be interpreted by taking into account that for negative β-values the probability of the event of interest occurring, is larger at point 0 than the corresponding probability at $x = 1$. Therefore, the design allocates more subjects to the experimental point where censoring is more likely.

The minimax D-optimal weights at point 0 for contamination functions in the classes $\mathcal{G}_1$ and $\mathcal{G}_2$ are found in Theorems 7 and 8 respectively. Figures 7.1 and 7.2 illustrate the behaviour of these optimal weights for various values of the constants c_1 and c_2 involved in the definitions of the classes of contaminants given in (7.3) and (7.4).

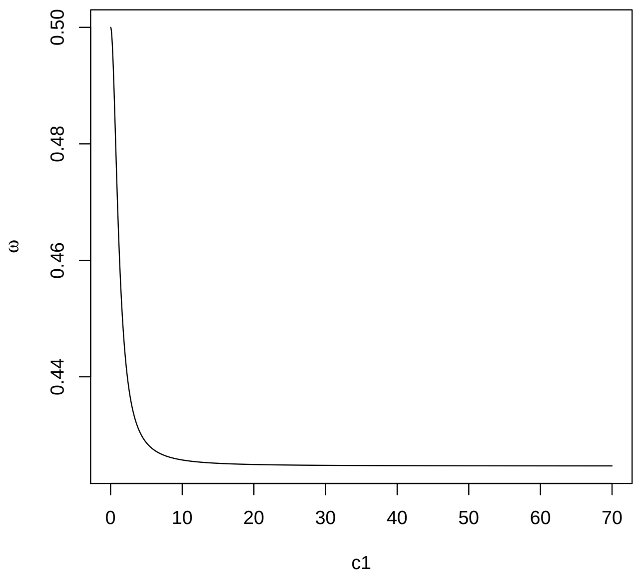


Figure 7.1: Minimax D-optimal weight ω at point 0 for $g \in \mathcal{G}_1$

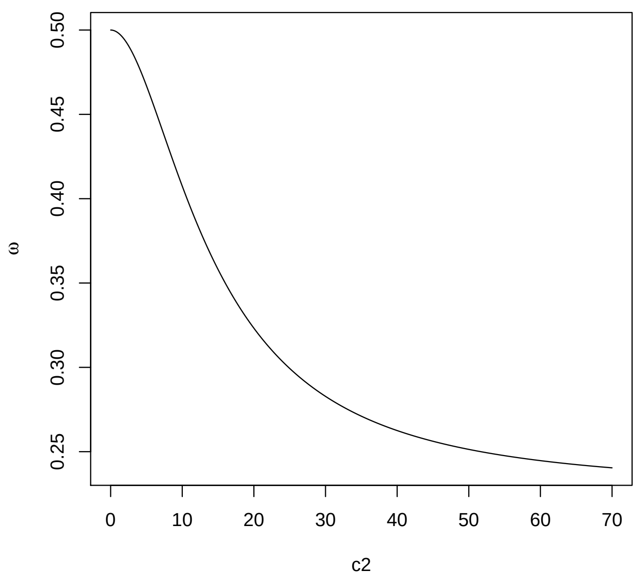


Figure 7.2: Minimax D-optimal weight ω at point 0 for $g \in \mathcal{G}_2$

We observe that in both cases of $g \in \mathcal{G}_1$ and $g \in \mathcal{G}_2$, the minimax D-optimal weight is smaller than 0.5 and its value further decreases with c_1 and c_2. Therefore, the balanced design allocating half the subjects at point 0 and the rest at point 1 is suboptimal for deviations from the exponential regression model.

The optimal weights appear to have limiting values as c_1 or c_2 increases. This means that if we allow the amount of contamination to increase the optimal design will not change much above certain values for c_1 and c_2.

In the case of the continuous design space $\mathcal{X} = [0, 1]$ and using any of the criteria defined in (7.6), (7.7) and (7.8) the support points of the resulting minimax designs are found to always be the points 0 and 1. However, the corresponding weights at point

$x = 0$ are the same as in the case of the binary design space $\mathcal{X} = \{0, 1\}$ described above. Hence even when the minimax design is supported at the boundaries of the continuous design space, the equal allocation rule leads to suboptimal designs.

7.6 Conclusions

In the case of a binary design space our results on minimax c-optimal designs for β, presented in section 7.3.1, coincide with the locally c-optimal designs for estimating β assuming the exponential regression model (see section 4.4 in Chapter 4). This means that the latter designs are also robust to departures from the exponential-based proportional hazards model provided the proportional hazards property is retained. Therefore, if one is interested in estimating only the covariate effect parameter β, the exponential distribution can be assumed for design construction without loss of generality.

If D-optimality is the desired criterion, that is, if estimation of both of the model parameters is required, then Theorems 7 and 8 provide analytical characterisations of the minimax optimal weights to be used. These characterisations, along with the numerical results we have for the Freireich data, suggest that we have to move away from the traditional balanced design to guard against misspecification of the exponential distribution.

We also provide analytical characterisations of minimax optimal weights based on either D- or c-optimality for an arbitrary continuous design space. Moreover, Theorems 9 and 10 offer a complete classification of the support points of the minimax c-optimal designs for estimating β and α respectively, for contamination functions in the class $\mathcal{G}_1$. Although similar analytical results could not be proven for the case of minimax D-optimal designs and $g \in \mathcal{G}_2$, our application to the Freireich data illustrates that the balanced design allocating equal proportions of subjects at the end-points of the design space will not perform well if the exponential regression model is incorrect. In particular, even if the minimax design has the boundaries of the design space as its support points, the corresponding weights will not be equal, thus making the balanced design suboptimal.

Chapter 8

Discussion

We now summarise the main results derived in this PhD thesis and discuss the importance and the benefits of our novel research. Furthermore, we propose future work that may be carried out in order to extent these results.

8.1 Results and conclusions

The research conducted in this project combines the known and well established criteria used in Design of Experiments with the features of the models that arise frequently in survival experiments and that are studied in Survival Analysis. This work is of great importance not only due to its direct applications in areas such as medicine, biostatistics, social sciences and engineering where survival and reliability models are met, but also because of the analytical methods proposed. Using designs based on DoE theory, we can minimise the cost and duration of survival experiments, while at the same time maximise the precision of their conclusions. The novelty of our research is clear from the small number of contributions to the literature in this area and, in particular, the lack of theoretical results on how experimental designs change in the presence of censoring, a phenomenon characterising survival data that arise in such experiments. Therefore, this work has a potential impact on how survival experiments are set up in the future. In particular our results are summarised as follows.

We have defined a wide class of two-parameter nonlinear models based on the form of, and some extra assumptions on, the information matrix for an arbitrary approximate design ξ. Our assumptions are satisfied by some of the most frequently used survival models and therefore the problem of constructing optimal designs is solved in more generality. Moreover, these assumptions can easily be verified and in particular they are easier to check than those of Yang and Stufken (2009). Therefore, our results can be directly applied by practitioners specifically working in survival experiments.

For models in the general class considered, we provide analytical characterisations of locally D- and c-optimal designs which can then be used as a starting point for the construction of parameter robust designs, for example, sequential designs. Based on some easily verifiable conditions on the model parameters, we completely classify locally D- and c-optimal designs and either solve the design problem entirely or simplify it, in terms of reducing the dimensions of the optimisation required for their construction.

The characterisation of locally c-optimal designs revealed that the weights corresponding to the two support points of the design must always be unequal for both a binary and a continuous design space. Therefore, if one is interested in estimating only the effect of the explanatory variable, which is often the primary aim of survival experiments, then different proportions of subjects should be allocated to the experimental points. This result is of great importance since practitioners traditionally plan survival experiments such as clinical trials using the equal allocation rule.

Through our application to the exponential-based proportional hazards model under Type-I and random censoring, we explicitly illustrate how censoring affects the optimal choice of design in the case of a continuous design space. In particular, for both censoring mechanisms the locally D-optimal designs maintain the allocation of equal numbers of subjects at the experimental points but these are not the end-points of the design space in all scenarios. When the event times are highly dependent on the explanatory variable under investigation, and therefore its effect is large, the design includes points where censoring is less likely to occur and hence are more informative. Depending on the sign of the β-parameter representing the covariate effect, the locally D-optimal design is always supported at the most informative boundary of the design space whereas the second support point tends to be away from the other design space end-point. On the other hand, even for small effect sizes, the locally c-optimal designs put more weight at the experimental point where the probability of censoring is greater in order for the information to be maximised. In either case, however, the standard design currently used in practice that puts the same proportion of observations at the two boundaries of the design space is suboptimal.

The parameter dependence of locally optimal designs makes them difficult to use in practical situations. In order to overcome this shortcoming we found parameter robust designs when there is some information about the parameter values. In particular, we assumed that the experimenter has a good approximation for the value of the parameter α, involved in the baseline hazard, from previous studies and that he/she can specify a range of β-values for the expected effect size. This is a situation that frequently arises in survival experiments and therefore our results can be implemented directly in practice.

Using this parameter information and for models in the general class we have defined, we constructed optimal designs based on standardised maximin criteria which maximise the worst efficiency among all two-point designs and therefore protect against the worst case scenario. Using our results on locally c-optimal designs, we have provided an analytical characterisation of standardised maximin c-optimal two-point designs that specifies the design entirely. As before we have shown that for c-optimality the commonly used equal allocation rule leads to a suboptimal design. Moreover, we have produced a complete classification of standardised maximin D-optimal two-point designs which reduces the numerical effort of design search substantially. To the best of our knowledge this is the first analytical characterisation of such designs when an explicit form for the locally D-optimal designs is not available.

Additionally, cluster designs were built since these are proven to be similarly effective and much more easily computed than other parameter-robust designs such as Bayesian optimal designs. For both the D- and c-optimality criteria we have facilitated their construction even more by our results on the corresponding locally optimal designs.

Using the exponential regression model in its proportional hazards parametrisation and a set of parameter values based on the well known survival data used in Freireich et al. (1963) we have shown that both parameter-robust design strategies are good alternatives to the locally optimal designs. If one is interested in using a design which, in the worst case scenario of misspecification of the model parameters, will have the best minimum efficiency, then the standardised maximin optimal designs are the best choice. Moreover, for D-optimality, that is, when both model parameters are to be estimated, cluster designs with more than two support points can also be used, thereby enabling any lack of fit of the model to be checked.

In practice survival data are often modelled through Cox's proportional hazards model instead of the corresponding parametric proportional hazards models since the former leaves the baseline hazard function unspecified and therefore involves fewer assumptions. We have met the needs of this practical scenario that requires the use of the partial likelihood function, introduced by Cox (1972), for the estimation of the model parameters by setting up a general framework for the construction of optimal designs for the Cox model. Our approach contains the results by Kalish and Harrington (1988) as a special case and it differs from that of López-Fidalgo and Rivas-López (2012) in that we work directly with the asymptotic covariance matrix, thereby avoiding the need for another level of approximation.

We have derived a general expression for the asymptotic covariance matrix for partial likelihood estimation, thus generalising the optimality criterion found in Kalish and Harrington (1988) for a binary design space. We illustrated our approach for the case of only one covariate and found a necessary condition for a design to be optimal for partial likelihood estimation. This can be used to discard candidate designs which do not satisfy this condition. This is a non-standard optimality criterion, and there is no such result in the literature yet.

Kalish and Harrington (1988) claim that the balanced design that allocates equal proportions of subjects to the support points of the design is sufficiently efficient under both Type-I and random censoring. However, we found this not to be the case for large effect sizes β and/or heavy censoring. We further extended a result by Kalish and Harrington (1988) to the case of a continuous design space. In particular, for Type-I censoring, the optimal design for partial likelihood estimation will not depend on the shape of the baseline hazard function, but only on the value of the survivor function at the censoring time c. This means that the optimal designs for constant baseline hazard, that is, for the exponential regression model, will be highly efficient for partial likelihood estimation in any underlying hazard model and can thus be regarded as optimal for all proportional hazards models.

Optimal designs for partial likelihood estimation are not trivial to find, and may therefore not be popular with practitioners. We have compared these designs with the locally c-optimal designs for the corresponding parametric model constructed using the full likelihood information, and found that the optimal designs for both methods are very similar, in particular, for heavy censoring which is often observed in practice. We used Taylor expansions to show that the two asymptotic variances are indeed approximately equal, where the accuracy of the approximation is higher, the heavier the censoring. Hence the c-optimal designs found in Chapter 4 for the general class of two-parametric nonlinear models are highly efficient also for partial likelihood estimation, and can thus be used without detriment in most situations.

When parametric models are used in practice, often the exponential distribution is naturally assumed for the times to event along with the proportional hazards assumption. However, this parametric model may hold only approximately. For this reason, we have defined a class of models in a neighbourhood of the exponential-based proportional hazards models. This class is specified by small deviations from the exponential distribution but nonetheless includes the next most frequently considered parametric proportional hazards models based on the Weibull and Gompertz distributions. Therefore, we take into account any imprecisions in the specification of the mean response that may occur in practice.

Following Wiens (1992), we use criteria based on the mean squared error matrix due to the bias of the maximum likelihood estimators for the model parameters. These correspond to the classical D- and c-optimality criteria and the minimax designs constructed minimise the corresponding criteria functions for the worst possible deviations from the exponential regression model within the class. We also incorporate Type-I censoring in the derivation of the mean squared error matrix and thus our resulting model robust designs also take into account the effect of censoring.

For both binary and continuous design spaces and for D- and c-optimality, we provide analytical characterisations for the optimal weights of the minimax designs. In particular, for a binary design space our results on minimax c-optimal designs for estimating the covariate coefficient β show that any deviations from the exponential distribution do not affect the optimal choice of design. Therefore, in survival experiments comparing, for example, two treatments or methods, where interest centres on estimating this treatment/method effect, the c-optimal design for the exponential-based proportional hazards model can be used without detriment. This again highlights the importance of our results from Chapters 4 and 5.

We finally note that our work presented in Chapters 4 and 5 on locally optimal and parameter robust designs led to the publication entitled "Optimal designs for two-parameter nonlinear models with application to survival models" in Statistica Sinica (Konstantinou, Biedermann and Kimber (2013)).

8.2 Future work

As mentioned in Chapter 5, a topic for further investigation is the analytical characterisation of standardised maximin D-optimal designs on a continuous design space in the case of positive values for the parameter β. In many clinical trial applications the response variable is the lifetime of patients and so larger event times are preferable. Since a new treatment will be studied if it is expected to increase the life expectancy of patients, the β-value will be negative. However, this is not the case for studies where the event of interest is non-fatal and so smaller event times may be desired. The method of proof we used to classify standardised maximin D-optimal designs in the case $\beta < 0$ does not work for the $\beta > 0$. Therefore, a different approach is needed in order to complete the classification of standardised maximin D-optimal designs.

For the Cox's model, the derivation of analytical results has proved to be difficult since the integrals involved in the maximum partial likelihood method cannot be solved analytically. However, based on our conclusions on the similarity of the variances and hence of the optimal designs found for full and partial likelihood estimation, we feel that a further theoretical investigation of these findings is worthwhile and may produce the long awaited analytical results for Cox's proportional hazards model.

Another possible future direction that can be explored is the construction of model-robust designs when the proportional hazards assumption is violated. An alternative class of models that can be used in this case is the class of accelerated failure time (AFT) models. Therefore, a similar approach to that used in Chapter 7, can be used to construct optimal designs that are robust to deviations from the exponential regression model but within an AFT framework rather than the proportional hazards framework we have used.

Our designs for Cox's model and the model-robust designs given in Chapter 7 are locally optimal with respect to the model parameter values. Hence, methods to make them robust to misspecifications of these values could also be topic for further investigation.

Finally, our results can be extended to models that involve more than one explanatory variable. Although this is not often encountered in survival experiments, such as clinical trials where only the treatment effect is investigated and all other factors affecting the response are not controlled by the experimenter, the incorporation of many explanatory variables that are under the experimenter's control is a natural extension to our existing findings.

Appendix A

A.1 Proofs for Chapter 1

A.1.1 Statement 1

Let $Q(\theta)$ be a positive function for all $\theta \in \mathbb{R}$, twice continuously differentiable and strictly increasing on $\mathbb{R}$ (conditions (a) and (b)). If the function $\log Q(\theta)$ is concave for $\theta \in \mathbb{R}$ (condition (d1)), then for any $s \in \mathbb{R}$, the function $g_2(\theta) = Q(\theta)(s-\theta)^2$ satisfies $g_2'(\theta) = 0$ for exactly two values of $\theta \in (-\infty, s]$ (condition (d)).

Proof. Since $Q(\theta)$ is twice continuously differentiable then $\log Q(\theta)$ is concave if and only if its second order derivative is non-positive for all $\theta \in \mathbb{R}$. That is, if the function $Q'(\theta)/Q(\theta)$ is decreasing on $\mathbb{R}$. Now

$$\frac{g_2'(\theta)}{(s-\theta)} = 0 \iff \frac{Q'(\theta)}{Q(\theta)} = \frac{2}{(s-\theta)}.$$

It can be easily seen that the right-hand side of this equation is an increasing function for $\theta \in (-\infty, s)$, whereas the left-hand side is decreasing on $(-\infty, s)$ using the definition of a concave function given above. Hence the equation $g_2'(\theta)/(s-\theta) = 0$ has exactly one solution on $(-\infty, s)$ which implies condition (d).

□

A.1.2 Statement 2

The GLM's with response variable following a Gamma$(\gamma, (k_1 + e^{\alpha+\beta x})^{k_2})$ or an Inv-Gamma$(\gamma, (k_1 + e^{\alpha+\beta x})^{k_2})$ distribution, where $\gamma > 0$, $k_1 > 0$, $k_2 \neq 0$ are known, have an information matrix of the form (4.1) and satisfy conditions (a)-(d) and (d1).

Proof. Let $T \sim$ Gamma$(\gamma, (k_1 + e^{\alpha+\beta x})^{k_2})$. The log-likelihood function at point x is then given by

$$l(x, \alpha, \beta) = \log f_G(t, x) = (\gamma - 1)\log t - \frac{t}{(k_1 + e^{\alpha+\beta x})^{k_2}} - \gamma k_2 \log\left(k_1 + e^{\alpha+\beta x}\right) - \log\Gamma(\gamma),$$

where $f_G(t,x)$ is the probability density function of the Gamma distribution and $\Gamma(\gamma)$ is the gamma function evaluated at the shape parameter γ. Therefore the second order derivatives of the log-likelihood function with respect to the parameters α and β are given by

$$\frac{\partial^2 l(x,\alpha,\beta)}{\partial\alpha^2} = \frac{tk_2 e^{\alpha+\beta x}(k_1 - k_2 e^{\alpha+\beta x})}{(k_1+e^{\alpha+\beta x})^{k_2+2}} - \frac{\gamma k_2 k_1 e^{\alpha+\beta x}}{(k_1+e^{\alpha+\beta x})^2},$$
$$\frac{\partial^2 l(x,\alpha,\beta)}{\partial\alpha\partial\beta} = x\frac{\partial^2 l(x,\alpha,\beta)}{\partial\alpha^2}, \quad \frac{\partial^2 l(x,\alpha,\beta)}{\partial\beta^2} = x^2\frac{\partial^2 l(x,\alpha,\beta)}{\partial\alpha^2}.$$

Using the fact that $E(T) = \gamma(k_1+e^{\alpha+\beta x})^{k_2}$, the Fisher information at point x is

$$I(x,\alpha,\beta) = \begin{pmatrix} E\left(-\frac{\partial^2 l}{\partial\alpha^2}\right) & E\left(-\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) \\ E\left(-\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) & E\left(-\frac{\partial^2 l}{\partial\beta^2}\right) \end{pmatrix} = \gamma\left(\frac{k_2 e^{\alpha+\beta x}}{k_1+e^{\alpha+\beta x}}\right)^2 \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix},$$

which yields (4.1) for $Q(\theta) = \gamma k_2^2 e^{2\theta}/(k_1+e^\theta)^2$.

If $T \sim$ Inv-Gamma$(\gamma, (k_1+e^{\alpha+\beta x})^{k_2})$ then the log-likelihood at x is

$$l(x,\alpha,\beta) = \gamma k_2 \log(k_1+e^{\alpha+\beta x}) - \frac{(k_1+e^{\alpha+\beta x})^{k_2}}{t} - (\gamma+1)\log t - \log\Gamma(\gamma),$$

and $E(1/T) = \gamma/(k_1+e^{\alpha+\beta x})^{k_2}$. Following the same procedure as before, we again obtain an information matrix of the form (4.1) for the same Q-function. It therefore remains to show that this Q-function satisfies assumptions (a)-(d) and (d1).

Since $\gamma > 0$, $k_1 > 0$ and $k_2 \neq 0$, the function $Q(\theta) = \gamma k_2^2 e^{2\theta}/(k_1+e^\theta)^2$ is positive for all $\theta \in \mathbb{R}$. Also

$$Q'(\theta) = 2\gamma k_1 k_2^2 e^{2\theta}/(k_1+e^\theta)^3 > 0 \ \forall\theta\in\mathbb{R}, \quad Q''(\theta) = 2\gamma k_1 k_2^2 e^{2\theta}(2k_1 - e^\theta)/(k_1+e^\theta)^4,$$

and so conditions (a) and (b) hold. Now

$$g_1''(\theta) = 4k_1(2k_1+e^\theta)/\gamma k_2^2 e^{2\theta}$$

is decreasing with θ, as its derivative with respect to θ is given by

$$-4k_1(4k_1+e^\theta)/\gamma k_2^2 e^{2\theta},$$

and therefore it is an injective function. This concludes the proof for condition (c). Moreover,

$$(\log Q(\theta))'' = -2k_1 e^\theta/(k_1+e^\theta)^2 < 0, \quad \forall\theta\in\mathbb{R}.$$

Hence condition (d1) and therefore condition (d) are also satisfied.

□

A.1.3 Statement 3

Any parametric proportional hazards model with hazard function of the form (4.2) and response variable subject to Type-I censoring, has information matrix of the form (4.1) and satisfies conditions (a)-(d) and (d1).

Proof. Let $Y_j = min\{T_j, c\}$, $j = 1, \ldots, n$ be random variables for the possibly right-censored observations and T_j follow model (4.2) with corresponding probability density and survivor functions given by

$$f(t_j) = e^{\alpha+\beta x_j} r(t_j) e^{-e^{\alpha+\beta x_j} \int_0^{t_j} r(s)\,ds}, \quad S(t_j) = e^{-e^{\alpha+\beta x_j} \int_0^{t_j} r(s)\,ds}, \qquad t_j > 0.$$

Also let the indicator variable

$$\delta_i = \begin{cases} 1, & \text{when } Y_j = T_j \\ 0, & \text{when } Y_j = c \end{cases},$$

that is, $\delta_j = 0$ when the jth observation is right-censored and unity otherwise. Then the log-likelihood function at point x_j is

$$l(x_j, \alpha, \beta) = \log\left(\{f(y_j)\}^{\delta_j}\{S(y_j)\}^{(1-\delta_j)}\right) = \delta_j[\alpha + \beta x_j + \log r(y_j)] - e^{\alpha+\beta x_j}\int_0^{y_j} r(s)ds.$$

In the second order derivatives of the above function with respect to α and β, the only random term involved is $\int_0^{y_j} r(s)ds$ with expectation given by

$$E\left(\int_0^{Y_j} r(s)ds\right) = \int_0^c \int_0^{y_j} r(s)\,ds\; f(y_j)\,dy_j + \int_0^c r(s)\,ds\; P(Y_j = c)$$

$$= \int_0^c \int_0^{y_j} r(s)\,ds\;\; e^{\alpha+\beta x_j}\; r(y_j)\; e^{-e^{\alpha+\beta x_j}\int_0^{y_j} r(s)\,ds}\, dy_j + \int_0^c r(s)\,ds\;\; e^{-e^{\alpha+\beta x_j}\int_0^c r(s)\,ds}$$

$$= \left(1 - e^{-e^{\alpha+\beta x_j}\int_0^c r(s)\,ds}\right) / e^{\alpha+\beta x_j}.$$

The resulting Fisher information matrix at point x_j

$$I(x, \alpha, \beta) = \begin{pmatrix} E\left(-\frac{\partial^2 l}{\partial\alpha^2}\right) & E\left(-\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) \\ E\left(-\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) & E\left(-\frac{\partial^2 l}{\partial\beta^2}\right) \end{pmatrix} = \left(1 - e^{-e^{\alpha+\beta x_j}\int_0^c r(s)\,ds}\right)\begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix},$$

is of the form (4.1) for $Q(\theta) = 1 - e^{-e^{\theta}\int_0^c r(s)\,ds}$.

From the parametrisation defined in (4.2) we observe that $r(s) > 0\; \forall s \in [0, c]$, since the baseline hazard function is always positive, and so $Q(\theta) = 1 - e^{-e^{\theta}\int_0^c r(s)\,ds} > 0$,

$\forall\theta \in \mathbb{R}$. The first and second order derivatives of $Q(\theta)$ are given by

$$Q'(\theta) = e^{\theta}\int_0^c r(s)\,ds\ e^{-e^{\theta}\int_0^c r(s)\,ds} > 0 \quad \forall\theta \in \mathbb{R}$$
$$Q''(\theta) = e^{\theta}\int_0^c r(s)\,ds\ e^{-e^{\theta}\int_0^c r(s)\,ds}\left(1 - e^{\theta}\int_0^c r(s)\,ds\right).$$

Hence conditions (a) and (b) are satisfied.

Condition (d1) is true if and only if

$$(\log Q(\theta))'' = \frac{Q''(\theta)Q(\theta) - \{Q'(\theta)\}^2}{\{Q(\theta)\}^2} \le 0$$
$$\iff Q''(\theta)Q(\theta) - \{Q'(\theta)\}^2 \le 0$$
$$\iff e^{\theta}\int_0^c r(s)\,ds\ e^{-e^{\theta}\int_0^c r(s)\,ds}\left(1 - e^{-e^{\theta}\int_0^c r(s)\,ds} - e^{\theta}\int_0^c r(s)\,ds\right) \le 0.$$

Now $1 - e^{-e^{\theta}\int_0^c r(s)\,ds} - e^{\theta}\int_0^c r(s)\,ds < 1 - \left(1 - e^{\theta}\int_0^c r(s)\,ds\right) - e^{\theta}\int_0^c r(s)\,ds = 0$. So (d1) and as a result condition (d) hold.

Let $\zeta = e^{\theta}\int_0^c r(s)\,ds$, which is positive for all $\theta \in \mathbb{R}$. The function $e^{\theta}\int_0^c r(s)\,ds \to \zeta$ is continuous and strictly increasing, therefore injective, and also surjective. Hence it is a bijective function. Using this re-parametrisation the second derivative of the function g_1, defined in condition (c), is given by

$$g_1''(\zeta) = \frac{-2\zeta e^{-\zeta}}{(1 - e^{-\zeta})^3}(1 - \zeta - e^{-\zeta} - \zeta e^{-\zeta}).$$

This is a strictly decreasing function for $\theta \in \mathbb{R}$ and therefore condition (c) is satisfied. In order to show this, it is enough to prove that $g_1'''(\zeta) < 0\ \forall\zeta > 0$ which is equivalent to showing

$$(1 - e^{-\zeta})^2 + \zeta(-3 + \zeta + 3e^{-2\zeta} + 4\zeta e^{-\zeta} + 4e^{-2\zeta}) > 0,\ \ \forall\zeta > 0.$$

A sufficient condition for this to be true is that $\rho_1(\zeta) := -3 + \zeta + 3e^{-2\zeta} + 4\zeta e^{-\zeta} + 4e^{-2\zeta}$ is strictly positive for all $\zeta > 0$. But $\rho_1(0) = 0$ and if $\rho_1(\zeta)$ is a strictly increasing function then $\rho_1(\zeta) > 0\ \forall\zeta > 0$. Now

$$\rho_1'(\zeta) > 0\ \ \forall\zeta > 0 \iff \rho_2(\zeta) := 1 - 5e^{-2\zeta} + 4e^{-\zeta} - 4\zeta e^{-\zeta} - 2\zeta e^{-2\zeta} > 0\ \ \forall\zeta > 0.$$

We observe that $\rho_2(0) = 0$ and so for $\rho_2(\zeta) > 0$ it is enough to show that

$$\rho_2'(\zeta) > 0\ \ \forall\zeta > 0 \iff 4e^{-\zeta}(2e^{-\zeta} - 2 + \zeta e^{-\zeta} + \zeta) > 0\ \ \forall\zeta > 0$$
$$\iff \rho_3(\zeta) := 2e^{-\zeta} - 2 + \zeta e^{-\zeta} + \zeta > 0\ \ \forall\zeta > 0.$$

This last statement is true because $\rho_3(0) = 0$ and $\rho_3'(\zeta) = 1 - e^{-\zeta} - \zeta e^{-\zeta} > 0 \ \forall \zeta > 0$ since $e^{\zeta} > 1 + \zeta \iff e^{-\zeta}(1+\zeta) < 1$.

□

A.1.4 Statement 4

Any parametric proportional hazards model with hazard function of the form (4.2), where $\int_0^{c_j} r(s)\,ds$ is log-concave in c_j on $\mathbb{R}^+$, and response variable subject to random censoring, has information matrix of the form (4.1) and satisfies conditions (a), (b), (d) and (d1). Condition (c) has to be checked on a case by case basis.

Proof. In the case of random censoring we assume that the subjects enter the experiment at random times $Z_j \in [0, c]$, $j = 1, \ldots, n$, which are independent of the times to event T_j and we also assume that $Z_j \sim U(0, c)$. Hence the censoring times $C_j = c - Z_j$ are also random and $C_j \sim U(0, c)$ with probability density function $f_{C_j}(c_j) = 1/c$. Under this censoring mechanism what we actually observe for each subject is $Y_j = min\{T_j, C_j\}$ with T_j following model (4.2). The log-likelihood at point x_j is

$$l(x_j, \alpha, \beta) = \delta_j(\log c + \alpha + \beta x_j + \log r(y_j)) + \log c - e^{\alpha+\beta x_j} \int_0^{y_j} r(s)\,ds,$$

where, as in the Type-I censoring case, $\delta_j = 0$ for a right-censored observation, that is $Y_j = C_j$ and $\delta_j = 1$ otherwise, that is $Y_j = T_j$. Now

$$E\left(\int_0^{Y_j} r(s)\,ds \middle| C_j = c_j\right) = \left(1 - e^{-e^{\alpha+\beta x_j}\int_0^{c_j} r(s)\,ds}\right) / e^{\alpha+\beta x_j},$$

and so

$$E\left(\int_0^{Y_j} r(s)\,ds\right) = E\left(E\left(\int_0^{Y_j} r(s)\,ds \middle| C_j = c_j\right)\right) = \frac{1}{c}\int_0^{c} \frac{1 - e^{-e^{\alpha+\beta x_j}\int_0^{c_j} r(s)\,ds}}{e^{\alpha+\beta x_j}}\,dc_j.$$

Hence the Fisher information matrix at point x_j is

$$I(x, \alpha, \beta) = \begin{pmatrix} E\left(-\frac{\partial^2 l}{\partial\alpha^2}\right) & E\left(-\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) \\ E\left(-\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) & E\left(-\frac{\partial^2 l}{\partial\beta^2}\right) \end{pmatrix} = \frac{1}{c}\int_0^{c} (1 - e^{-e^{\alpha+\beta x_j}\int_0^{c_j} r(s)\,ds})\,dc_j \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix},$$

which yields (4.1) for $Q(\theta) = \frac{1}{c}\int_0^{c} (1 - e^{-e^{\theta}\int_0^{c_j} r(s)\,ds})\,dc_j$.

For fixed c_j, $1 - e^{-e^{\theta}\int_0^{c_j} r(s)\,ds} > 0 \ \forall \theta \in \mathbb{R}$ since $r(s) > 0$. Therefore, conditions (a) and (b) are satisfied for the Q-function given above with first and second order derivatives given by

$$Q'(\theta) = \frac{1}{c}\int_0^c e^\theta \int_0^{c_j} r(s)\,ds\, e^{-e^\theta \int_0^{c_j} r(s)\,ds}\; dc_j > 0 \quad \forall \theta \in \mathbb{R}$$

$$Q''(\theta) = \frac{1}{c}\int_0^c e^\theta \int_0^{c_j} r(s)\,ds\, e^{-e^\theta \int_0^{c_j} r(s)\,ds}\left(1 - e^\theta \int_0^{c_j} r(s)\,ds\right)\; dc_j.$$

Condition (d1) is equivalent to the function $Q(\theta)$ being log-concave on $\mathbb{R}$. From Theorem 2 in Prékopa (1973) it follows that if the function $Q(\theta, c_j) = 1 - e^{-e^\theta \int_0^{c_j} r(s)\,ds}$ is log-concave then $\int_0^c 1 - e^{-e^\theta \int_0^{c_j} r(s)\,ds}\, dc_j$ is also log-concave and therefore condition (d1) is satisfied.

Now for the two variable function $Q(\theta, c_j)$ to be log-concave we must show that the Hessian matrix involving the second order derivatives of $\log Q(\theta, c_j)$ with respect to θ and c_j is negative semidefinite. This is true if and only if its diagonal entries are both non-positive and its determinant is non-negative.

From the proof of condition (d1) for Type-I censoring (section A.1.3), we have that $(\partial^2/\partial\theta^2)\log Q(\theta, c_j) < 0$. Also the determinant of the Hessian is non-negative if and only if

$$\begin{aligned}
&\left(1 - e^{-e^\theta \int_0^{c_j} r(s)\,ds} - e^\theta \int_0^{c_j} r(s)\,ds\right)\left(1 - e^{-e^\theta \int_0^{c_j} r(s)\,ds}\right) \\
&\left((\partial/\partial c_j) r(c_j) \int_0^{c_j} r(s)\,ds - \{r(c_j)\}^2\right) \geq 0 \\
\iff\;& (\partial/\partial c_j) r(c_j) \int_0^{c_j} r(s)\,ds - \{r(c_j)\}^2 \leq 0 \\
\iff\;& \frac{(\partial/\partial c_j) r(c_j) \int_0^{c_j} r(s)\,ds - \{r(c_j)\}^2}{\left(\int_0^{c_j} r(s)\,ds\right)^2} \leq 0 \\
\iff\;& \log \int_0^{c_j} r(s)\,ds \quad \text{is concave on } \mathbb{R}^+ \\
\iff\;& \int_0^{c_j} r(s)\,ds \quad \text{is log-concave on } \mathbb{R}^+,
\end{aligned}$$

which is true. Moreover for the second diagonal element $(\partial^2/\partial c_j^2)\log Q(\theta, c_j)$ to be non-positive we must show that

$$\frac{\partial r(c_j)}{\partial c_j)} r(c_j)\left(1 - e^{-e^\theta \int_0^{c_j} r(s)\,ds}\right) - e^\theta \{r(c_j)\}^2 \leq 0.$$

Using the assumption of $\int_0^{c_j} r(s)\,ds$ being log-concave on $\mathbb{R}^+$ it is enough to show that

$$\frac{\left(1 - e^{-e^\theta \int_0^{c_j} r(s)\,ds}\right)}{\int_0^{c_j} r(s)\,ds} - e^\theta \leq 0$$

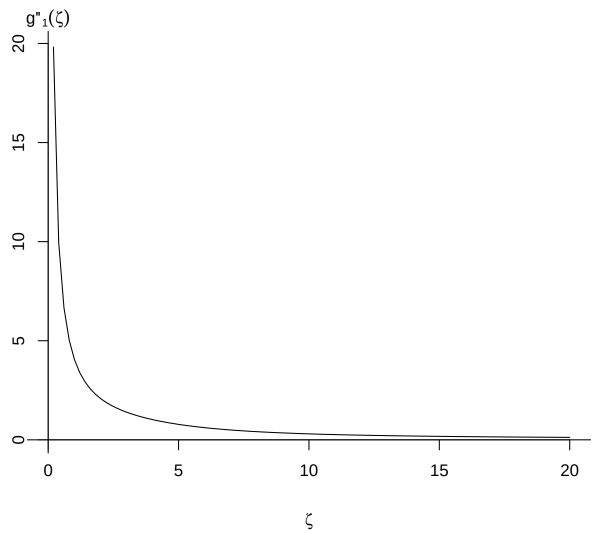
$$\iff 1 - e^{-e^{\theta}\int_0^{c_j} r(s)\,ds} - e^{\theta}\int_0^{c_j} r(s)\,ds \le 0,$$

which is true and therefore assumption (d1) is satisfied.

Condition (c) has to be checked on a case by case basis. For example, the Q-function for the exponential based proportional hazards model is $Q(\theta) = \left(ce^{\theta} - 1 + e^{-ce^{\theta}}\right)/ce^{\theta}$ and using the parametrisation $\zeta = ce^{\theta} > 0$ we have

$$g_1''(\zeta) = \frac{-2\zeta}{(\zeta - 1 + e^{-\zeta})^3}\left\{(\zeta^2 e^{-\zeta} - 1 + e^{-\zeta} + \zeta e^{-\zeta})(\zeta - 1 + e^{-\zeta}) - 2(1 - e^{-\zeta} - \zeta e^{-\zeta})^2\right\}.$$

From the graph given below we observe that this is a decreasing function for $\zeta > 0$ and therefore $g_1''(\theta)$ is decreasing on $\mathbb{R}$ which implies condition (c).



□

A.1.5 Statement 5

The accelerated failure time model with failure time distribution Gamma$(2, 1 + e^{\alpha+\beta x})$ for $\alpha + \beta x \le 0$ and Exponential with rate parameter equal to $\left(-2e^{(\alpha+\beta x)/2} + \frac{\alpha+\beta x}{2}\right)^{-1}$ for $\alpha + \beta x > 0$, has information matrix of the form (4.1), satisfies conditions (a)-(d) and (d1) and is not three times continuously differentiable.

Proof. As shown in section A.1.2, a Gamma$(2, 1 + e^{\alpha+\beta x})$ distribution yields (4.1) with corresponding Q-function $2e^{2\theta}/(1+e^{\theta})^2$. Now the log-likelihood function at point x for the Exponential distribution with rate parameter $\left(-2e^{(\alpha+\beta x)/2} + \frac{\alpha+\beta x}{2}\right)^{-1}$ is

$$l(x, \alpha, \beta) = \log\left(\frac{\alpha + \beta x}{2} - 2e^{(\alpha+\beta x)/2}\right) + \frac{t}{2e^{(\alpha+\beta x)/2} - \frac{\alpha+\beta x}{2}},$$

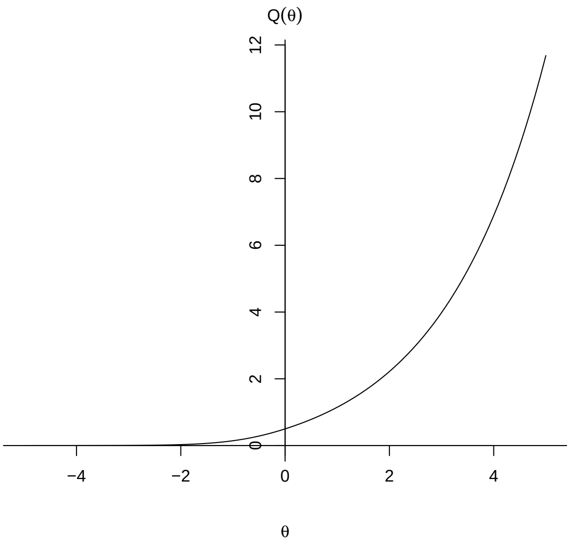
and so the Fisher information matrix at x is given by

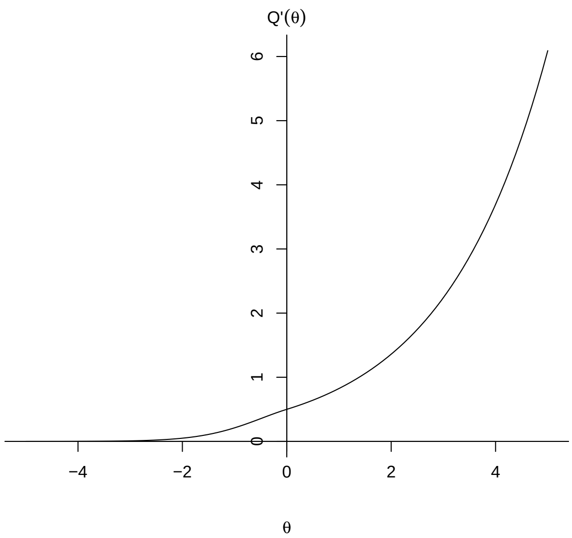
$$I(x, \alpha, \beta) = \left(e^{(\alpha+\beta x)/2} - \frac{1}{2}\right) \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix},$$

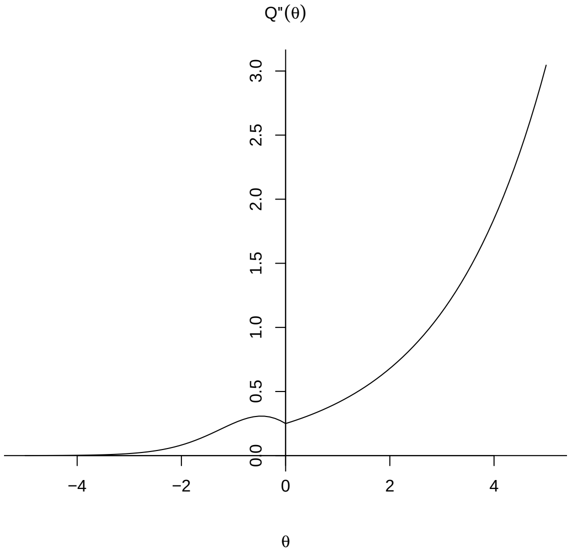
which yields (4.1) with Q-function $e^{\theta/2} - 1/2$. Therefore, the accelerated failure time model has an information matrix of the form (4.1) with

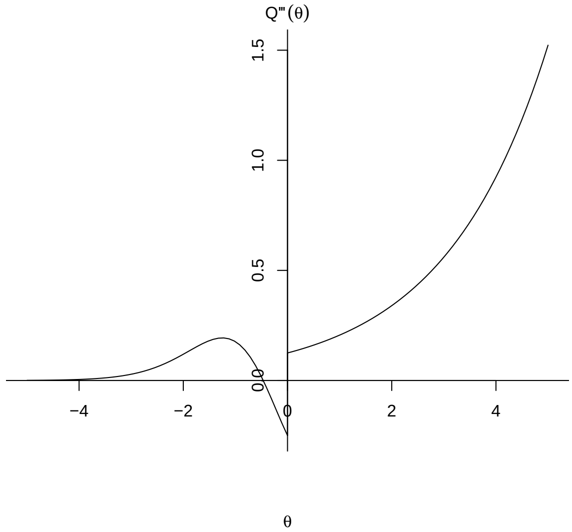
$$Q(\theta) = \begin{cases} 2e^{2\theta}/(1+e^{\theta})^2, & \text{if } \theta \leq 0 \\ e^{\theta/2} - 1/2, & \text{if } \theta > 0 \end{cases}.$$

The plots given below verify that conditions (a) and (b) hold but $Q(\theta)$ is not three times differentiable as continuity is disrupted at point $\theta = 0$.

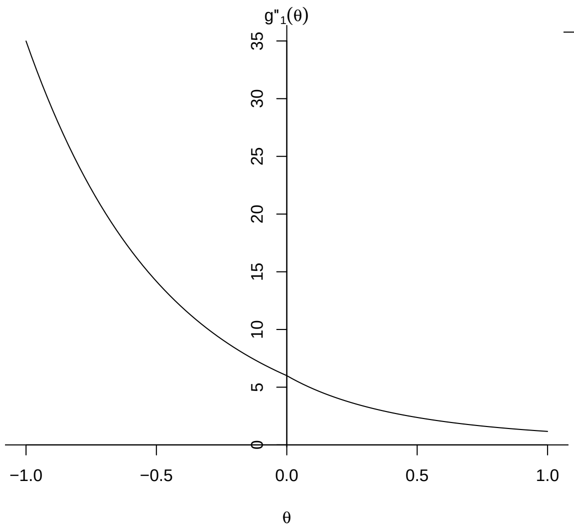


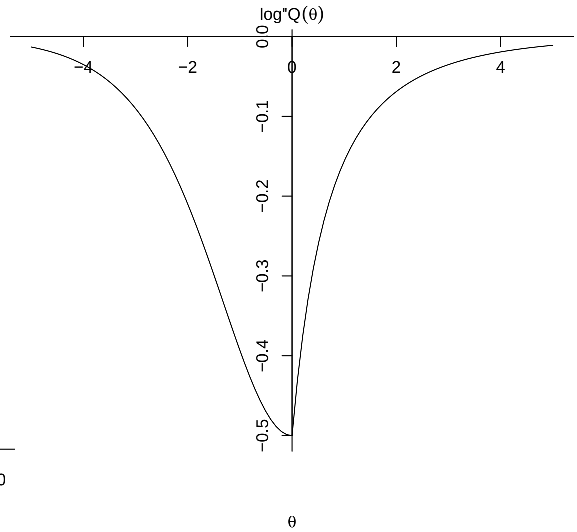






Moreover, as shown in the following two plots, the function $g_1''(\theta)$ defined in condition (c) is strictly decreasing and $(\log Q(\theta))'' < 0$ for all $\theta \in \mathbb{R}$ and therefore conditions (c) and (d1), and hence condition (d), are satisfied.





Hence assumptions (a)-(d) and (d1) hold for the accelerated failure time model considered but Yang and Stufken (2009) assumption of a three times continuously differentiable Q-function is not satisfied.

□

A.1.6 Lemma 1

Let $\beta \neq 0$ and conditions (a)-(c) be satisfied. Then the locally D-optimal design for a model with information matrix (4.1) is unique and has two equally weighted support points.

Proof. Let α and $\beta > 0$ be fixed. The case where $\beta < 0$ can be shown analogously and is therefore omitted. Also let ξ^* be a locally D-optimal design for a model with information matrix (4.1) and

$$M^{-1}(\xi^*, \alpha, \beta) = \begin{pmatrix} m_1 & m_2 \\ m_2 & m_3 \end{pmatrix},$$

where $m_1, m_2, m_3 \in \mathbb{R}$. From Theorem 1 in section 4.2.1 we obtain that a D-optimal design ξ^* must satisfy the inequality

$$m_1 + 2m_2 + m_3x^2 \leq 2/Q(\alpha + \beta x) \quad \forall x \in [0, 1],$$

with equality at the support points of ξ^*. Using the parametrisation $\theta = \alpha + \beta x$ this is equivalent to ξ^* satisfying the inequality

$$d(\theta) := d_1 + d_2\theta + d_3\theta^2 \leq 2/Q(\theta) = g_1(\theta) \quad \forall\theta \in [\alpha, \alpha + \beta],$$

with equality at the support points θ_i of ξ^*, where $d_1, d_2, d_3 \in R$.

Now suppose a locally D-optimal design has three support points, $\alpha \leq \theta_1 < \theta_2 < \theta_3 \leq \alpha + \beta$. Then $d(\theta_i) = g_1(\theta_i)$, $i = 1, 2, 3$. By Cauchy's mean value theorem, there exist points $\tilde{\theta}_i$, $i = 1, 2$ such that

$$\theta_1 < \tilde{\theta}_1 < \theta_2 < \tilde{\theta}_2 < \theta_3 \text{ and } d'(\tilde{\theta}_i) = g_1'(\tilde{\theta}_i).$$

Since $d(\theta) \leq g_1(\theta)$ on $[\alpha, \alpha + \beta]$, we also have $d'(\theta_2) = g_1'(\theta_2)$. By the mean value theorem, there exist points $\hat{\theta}_i$, $i = 1, 2$ such that

$$\tilde{\theta}_1 < \hat{\theta}_1 < \theta_2 < \hat{\theta}_2 < \tilde{\theta}_2 \text{ and } d''(\hat{\theta}_i) = g_1''(\hat{\theta}_i).$$

Now $d''(\theta)$ is constant and using condition (c) it can intersect with $g_1''(\theta)$ at most once on $[\alpha, \alpha + \beta]$, which contradicts the assumption of three support points. Hence a D-optimal design has exactly two support points, with equal weights.

Let ξ_1^* and ξ_2^* be two locally D-optimal designs. By log-concavity of the D-criterion, the design $\xi_3 = 0.5\xi_1^* + 0.5\xi_2^*$ must also be locally D-optimal. However, if ξ_1^* and ξ_2^* are different, ξ_3 has more than two support points, which contradicts the result above. Hence the locally D-optimal design ξ^* is unique.

□

A.1.7 Theorem 2(b)

Let conditions (a)-(d) be satisfied. If $\beta < 0$, the design

$$\xi^* = \begin{Bmatrix} 0 & x_1^* \\ 0.5 & 0.5 \end{Bmatrix},$$

is locally D-optimal on $\mathcal{X} = [0, 1]$, where $x_1^* = 1$ if $\beta > -2Q(\alpha + \beta)/Q'(\alpha + \beta)$. Otherwise, x_1^* is the unique solution of the equation $\beta x_1 + 2Q(\alpha + \beta x_1)/Q'(\alpha + \beta x_1) = 0$.

Proof. Let $\beta < 0$. From Lemma 1 we know that the locally D-optimal design is supported at exactly two points with equal weights. Let $x_0^*, x_1^* \in [0, 1]$ be the two equally weighted support points with $x_0^* < x_1^*$. Then the determinant of the information matrix (4.1) is given by

$$|M(\xi, \alpha, \beta)| = \frac{1}{4}Q(\alpha + \beta x_0^*)Q(\alpha + \beta x_1^*)(x_0^* - x_1^*)^2.$$

For fixed x_1^*,

$$\frac{\partial|M(\xi,\alpha,\beta)|}{\partial x_0^*)} = \frac{Q(\alpha+\beta x_1^*)(x_0^*-x_1^*)}{4}\left[\beta Q'(\alpha+\beta x_0^*)(x_0^*-x_1^*)+2Q(\alpha+\beta x_0^*)\right] < 0,$$

using conditions (a) and (b). Hence, regardless of the value of x_1^*, the determinant is decreasing with x_0^* and therefore maximised for $x_0^* = 0$. It remains to maximise

$$g_2(\alpha+\beta x_1) = Q(\alpha+\beta x_1)x_1^2.$$

Using assumption (d), $g_2(\alpha+\beta x_1)$ has exactly two turning points on $[0,\infty)$ which can be found by solving the equation

$$\beta Q'(\alpha+\beta x_1)x_1^2 + 2Q(\alpha+\beta x_1)x_1 = 0.$$

It is easy to see that one of the turning points is $x_1 = 0$ which is always a minimum, since

$$\left.\frac{\partial^2 g_2(\alpha+\beta x_1)}{\partial x_1^2}\right|_{x_1=0} = 2Q(\alpha) > 0,$$

and hence the other turning point must be a maximum. Now if this maximum is attained outside the design space $\mathcal{X} = [0,1]$, $g_2(\alpha+\beta x_1)$ is maximised at $x_1 = 1$ which will then be the second support point of the locally D-optimal design. This occurs if and only

$$\left.\frac{\partial g_2(\alpha+\beta x_1)}{\partial x_1}\right|_{x_1=1} > 0,$$

which is equivalent to

$$\beta > -Q(\alpha+\beta)/Q'(\alpha+\beta).$$

Otherwise the point at which the maximum is attained will be the larger support point x_1^*. This is found by solving

$$\frac{\partial g_2(\alpha+\beta x_1)}{\partial x_1} = 0,$$

which is equivalent to solving

$$\beta x_1 + 2Q(\alpha+\beta x_1)/Q'(\alpha+\beta x_1) = 0.$$

□

A.1.8 Theorem 4(b)

Let conditions (a), (b) and (d1) be satisfied. If $\beta < 0$, the design ξ^* with support points u and x_1^* and the optimal weights given in (4.4) is locally c-optimal for β on

$\mathcal{X} = [u, v]$, where $x_1^* = v$ if

$$\beta(u - v) - 2\frac{Q(\alpha + \beta v)}{Q'(\alpha + \beta v)}\left(1 + \frac{\sqrt{Q(\alpha + \beta v)}}{\sqrt{Q(\alpha + \beta u)}}\right) < 0.$$

Otherwise, x_1^* is the unique solution of the equation

$$\beta(u - x_1) - 2\frac{Q(\alpha + \beta x_1)}{Q'(\alpha + \beta x_1)}\left(1 + \frac{\sqrt{Q(\alpha + \beta x_1)}}{\sqrt{Q(\alpha + \beta u)}}\right) = 0.$$

Proof. Let $\beta < 0$ and using Lemma 2 let $x_0^* < x_1^*$ be the two support points of a locally c-optimal design for β. For this design with corresponding optimal weights given in (4.4) the objective function we want to minimise is given by

$$\tilde{d}(x_0^*, x_1^*) := \left(\frac{1}{\sqrt{Q(\alpha + \beta x_0^*)}} + \frac{1}{\sqrt{Q(\alpha + \beta x_1^*)}}\right)^2 \frac{1}{(x_0^* - x_1^*)^2}.$$

For fixed x_1^*,

$$\begin{aligned}\frac{\partial \tilde{d}(x_0^*, x_1^*)}{\partial x_0^*} &= \left(\frac{1}{\sqrt{Q(\alpha + \beta x_0^*)}} + \frac{1}{\sqrt{Q(\alpha + \beta x_1^*)}}\right)\frac{1}{(x_0^* - x_1^*)^2} \\ &\quad \left[\frac{-\beta Q'(\alpha + \beta x_0^*)}{\{Q(\alpha + \beta x_0^*)\}^{3/2}} - \frac{2}{(x_0^* - x_1^*)}\left(\frac{1}{\sqrt{Q(\alpha + \beta x_0^*)}} + \frac{1}{\sqrt{Q(\alpha + \beta x_1^*)}}\right)\right] > 0,\end{aligned}$$

using conditions (a) and (b). Hence regardless of the value of x_1^* $\tilde{d}(x_0^*, x_1^*)$ is increasing with x_0^* and therefore attains its minimum in $[u, v]$ at the lower bound, u, of the design space. It remains to minimise the function $\tilde{d}(u, x_1)$, the turning points of which can be found by solving $(\partial/\partial x_1)\tilde{d}(u, x_1) = 0$. This is equivalent to solving

$$u - x_1 = \frac{2Q(\alpha + \beta x_1)}{\beta Q'(\alpha + \beta x_1)}\left(1 + \frac{\sqrt{Q(\alpha + \beta x_1)}}{\sqrt{Q(\alpha + \beta u)}}\right).$$

We observe that $u - x_1$ is decreasing with x_1 and using conditions (a), (b) and (d1) the left-hand side is increasing with x_1 as the product of two positive and increasing functions. Hence the above equation has exactly one root and so $\tilde{d}(u, x_1)$ has exactly one turning point in (u, ∞). This must be a minimum as

$$\lim_{x_0 \to \infty} \tilde{d}(u, x_1) = \lim_{x_0 \to u} \tilde{d}(u, x_1) = \infty.$$

If the minimum is not in the interior of the design space, its upper bound, v, is the larger support point x_1^* of the locally c-optimal design for β. This occurs if and only if

$$\left.\frac{\partial \tilde{d}(u,x_1)}{\partial x_1}\right|_{x_1=v} < 0$$

which is equivalent to

$$\beta(u-v) - 2\frac{Q(\alpha+\beta v)}{Q'(\alpha+\beta v)}\left(1+\frac{\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}}\right) < 0.$$

Otherwise, x_1^* is the unique solution of

$$\frac{\partial \tilde{d}(u,x_1)}{\partial x_1}\tilde{d}(u,x_1) = 0,$$

which is equivalent to solving

$$\beta(u-x_1) - 2\frac{Q(\alpha+\beta x_1)}{Q'(\alpha+\beta x_1)}\left(1+\frac{\sqrt{Q(\alpha+\beta x_1)}}{\sqrt{Q(\alpha+\beta u)}}\right) = 0.$$

□

Appendix B

B.1 Proofs for Chapter 5

B.1.1 Statement 1 (Proof of Theorem 5)

Let

$$w(\beta) := \beta + \frac{2Q(\alpha+\beta)}{Q'(\alpha+\beta)},$$

with $w(\beta_0) \le 0$ and $u(x,\beta) := Q(\alpha+\beta x)x^2/Q(\alpha+\beta x_\beta)x_\beta^2$, where x_β satisfies the equation

$$\beta x_\beta + \frac{2Q(\alpha+\beta x_\beta)}{Q'(\alpha+\beta x_\beta)} = 0.$$

For fixed $0 < x \le 1$, the function $\beta \to u(x,\beta)$ is unimodal.

Proof. Using condition (d1) it can be easily shown that the function $w(\beta)$ is increasing with β. We consider two cases: (i) $w(\beta_1) > 0$ and (ii) $w(\beta_1) \le 0$.

Case (i): Since $w(\beta_0) \le 0$ and $w(\beta)$ is continuous there exists $\beta^* \in (\beta_0, \beta_1]$ such that $w(\beta) > 0$ for all $\beta \ge \beta^*$. In this case $x_\beta = 1$ and for fixed $0 < x \le 1$

$$\frac{\partial u(x,\beta)}{\partial\beta} = \frac{x^2}{[Q(\alpha+\beta)]^2}\left\{Q'(\alpha+\beta x)xQ(\alpha+\beta) - Q(\alpha+\beta x)Q'(\alpha+\beta)]\right\}.$$

From condition (d1), $Q'(\theta)/Q(\theta)$ is decreasing with θ and therefore

$$\frac{Q'(\alpha+\beta x)}{Q(\alpha+\beta x)}x \le \frac{Q'(\alpha+\beta x)}{Q(\alpha+\beta x)} \le \frac{Q'(\alpha+\beta)}{Q(\alpha+\beta)},$$

since $x \le 1$ ($\Rightarrow \alpha+\beta x \ge \alpha+\beta$). Hence the derivative of $u(x,\beta)$ with respect to β is non-positive for all $\beta \in [\beta^*, \beta_1]$ and $u(x,\beta)$ is minimised at β_1.

For $\beta < \beta^*$ and fixed $0 < x \le 1$, solving

$$\frac{\partial u(x,\beta)}{\partial\beta} = 0$$

is equivalent to solving

$$Q'(\alpha+\beta x)xQ(\alpha+\beta x_\beta)x_\beta - Q(\alpha+\beta x)\Big[Q'(\alpha+\beta x_\beta)x_\beta\left(x_\beta+\beta\frac{dx_\beta}{d\beta}\right)$$
$$+2Q(\alpha+\beta x_\beta)\frac{dx_\beta}{d\beta}\Big]=0. \tag{B.1}$$

Using the fact that

$$\beta x_\beta + \frac{2Q(\alpha+\beta x_\beta)}{Q'(\alpha+\beta x_\beta)} = 0$$

and substituting this expression for $Q(\alpha+\beta x_\beta)$, equation (B.1) becomes

$$\beta x + \frac{2Q(\alpha+\beta x)}{Q'(\alpha+\beta x)} = 0,$$

which has a unique solution β such that $x_\beta = x$ using part (b) of Theorem 2 (see section 4.2.3). Therefore the function $\beta \to u(x,\beta)$ is unimodal for fixed x.

Case (ii): If $w(\beta_1) \le 0$ then for all $\beta \in [\beta_0,\beta_1]$ $w(\beta) \le 0$ and following the same arguments as in the $\beta < \beta^*$ case, for fixed $0 < x \le 1$ the function $\beta \to u(x,\beta)$ is unimodal.

□

B.1.2 Statement 2 (Proof of Theorem 5)

The standardised maximin D-optimal two-point design $\xi^*_{\{0,x\}}$ is in the set

$$M_= := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) = u(x,\beta_1)\right\},$$

where

$$u(x,\beta) := \frac{Q(\alpha+\beta x)x^2}{Q(\alpha+\beta x_\beta)x_\beta^2}.$$

Proof. Let us assume that the design is in $M_< := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) < u(x,\beta_1)\right\}$. and so we must maximise the function $u(x,\beta_0)$. Taking its first derivative with respect to x and equating it to zero yields

$$\beta_0 x + \frac{2Q(\alpha+\beta_0 x)}{Q'(\alpha+\beta_0 x)} = 0,$$

which has a unique solution $x = x_{\beta_0}$. Hence $\{u(x_{\beta_0},\beta_0)\}^{1/2} = 1 < \{u(x_{\beta_0},\beta_1)\}^{1/2}$, which is a contradiction since the efficiency is always less than or equal to 1. Following similar arguments for set $M_> := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) > u(x,\beta_1)\right\}$ also leads to a contradiction and therefore the standardised maximin D-optimal two-point design must be in the $M_=$ set. □

B.2 Proofs for Chapter 6

B.2.1 Proposition 1

Let $\mathcal{H}$ be the class of all one-point designs where the support point is in the design space $\mathcal{X} = [u, v]$, and let $\eta = \{x; 1\} \in \mathcal{H}$. If a design ξ^* on $\mathcal{X}$ with support points $\{x_1, \ldots, x_m\}$ and corresponding weights $\{\omega_1, \ldots, \omega_m\}$ is optimal for estimating β via the partial likelihood method, the inequality

$$d(\xi^*, \eta) \leq 0$$

holds for all $\eta \in \mathcal{H}$, with equality in the one-point designs $\xi_i = \{x_i; 1\}$, $i = 1, \ldots, m$, generated by the support points of ξ^*. Here $d(\xi^*, \eta)$ is the Fréchet derivative of the criterion function at ξ^* in direction of the one-point design η, and is given by

$$\begin{aligned} d(\xi^*, \eta) &= -\sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q e^{\beta(x_i + x_q)} (x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{\sum_{l=1}^m \omega_l \pi_l(y) e^{\beta x_l}} dy \\ &\quad - \sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q e^{\beta(x_i + x_q)} (x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)\pi_x(y) e^{\beta x}}{(\sum_{l=1}^m \omega_l \pi_l(y) e^{\beta x_l})^2} dy \\ &\quad + \sum_{q=1}^{m} \omega_q e^{\beta(x + x_q)} (x - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_x(y)\pi_q(y)}{\sum_{l=1}^m \omega_l \pi_l(y) e^{\beta x_l}} dy, \end{aligned}$$

where $\pi_x(y)$ is the probability of being at risk at time y given covariate value x.

Proof. We first find the Fréchet derivative of the criterion function $\Sigma(\xi)$ defined in (6.5), for the case of one covariate, at a design ξ in the direction of another design η, where

$$\xi = \begin{Bmatrix} x_1 & \ldots & x_m \\ \omega_1 & \ldots & \omega_m \end{Bmatrix} \quad \text{and} \quad \eta = \begin{Bmatrix} x_{m+1} & \ldots & x_l \\ \omega_{m+1} & \ldots & \omega_l \end{Bmatrix}.$$

Then

$$(1-\varepsilon)\xi + \varepsilon\eta = \begin{Bmatrix} x_1 & \ldots & x_m & x_{m+1} & \ldots & x_l \\ \omega_1^* & \ldots & \omega_m^* & \omega_{m+1}^* & \ldots & \omega_l^* \end{Bmatrix}$$

where $\omega_i^* = (1-\varepsilon)\omega_i$ if $i \leq m$ or $\omega_i^* = \varepsilon\omega_i$ if $i > m$. Let $R_1(y) = \sum_{r=1}^m \omega_r \pi_r(y) \exp(\beta x_r)$ and $R_2(y) = \sum_{r=m+1}^l \omega_r \pi_r(m) \exp(\beta x_r)$. Then

$$\begin{aligned} &\Sigma((1-\varepsilon)\xi + \varepsilon\eta) - \Sigma(\xi) \\ &= \sum_{i=1}^{l} \sum_{q<i} \omega_i^* \omega_q^* \exp(\beta(x_i + x_q))(x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{(1-\varepsilon)R_1(y) + \varepsilon R_2(y)} dy \\ &\quad - \sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q \exp(\beta(x_i + x_q))(x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{R_1(y)} dy \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q \exp(\beta(x_i + x_q))(x_i - x_q)^2 \\
&\quad \int_0^\infty h_0(y)\pi_i(y)\pi_q(y) \left[\frac{(1-\varepsilon)^2}{(1-\varepsilon)R_1(y) + \varepsilon R_2(y)} - \frac{1}{R_1(y)} \right] dy \\
&\quad + (1-\varepsilon)\varepsilon \sum_{i=m+1}^{l} \sum_{q=1}^{m} \omega_i \omega_q \exp(\beta(x_i + x_q))(x_i - x_q)^2 \\
&\quad \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{(1-\varepsilon)R_1(y) + \varepsilon R_2(y)} \, dy + O(\varepsilon^2) \\
&= \sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q \exp(\beta(x_i + x_q))(x_i - x_q)^2 \\
&\quad \int_0^\infty h_0(y)\pi_i(y)\pi_q(y) \frac{-\varepsilon(R_1(y) + R_2(y)) + O(\varepsilon^2)}{R_1(y)[(1-\varepsilon)R_1(y) + \varepsilon R_2(y)]} \, dy \\
&\quad + \varepsilon \sum_{i=m+1}^{l} \sum_{q=1}^{m} \omega_i \omega_q \exp(\beta(x_i + x_q))(x_i - x_q)^2 \\
&\quad \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{(1-\varepsilon)R_1(y) + \varepsilon R_2(y)} \, dy + O(\varepsilon^2).
\end{aligned}
$$

The Fréchet derivative is therefore

$$
\begin{aligned}
d(\xi, \eta) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} (\Sigma((1-\varepsilon)\xi + \varepsilon\eta) - \Sigma(\xi)) \\
&= -\sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q \exp(\beta(x_i + x_q))(x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{R_1(y)} \, dy \\
&\quad - \sum_{i=1}^{m} \sum_{q<i} \omega_i \omega_q \exp(\beta(x_i + x_q))(x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)R_2(y)}{R_1^2(y)} \, dy \\
&\quad + \sum_{i=m+1}^{l} \sum_{q=1}^{m} \omega_i \omega_q \exp(\beta(x_i + x_q))(x_i - x_q)^2 \int_0^\infty \frac{h_0(y)\pi_i(y)\pi_q(y)}{R_1(y)} \, dt.
\end{aligned}
$$

Clearly, $d(\xi, \eta) = \sum_{i=m+1}^{l} \omega_i d(\xi, \eta_i)$, where η_i is the one-point design with support x_i and weight 1, $i = m+1, \ldots, l$. (Equivalently, it can be shown that the Gâteaux derivative is linear in its second argument.) Therefore we only need to consider directions towards one-point designs. If ξ is optimal, $\Sigma((1-\varepsilon)\xi + \varepsilon\eta_i) - \Sigma(\xi) \le 0$ for all designs $\eta_i \in \mathcal{H}$, and the inequality $d(\xi^*, \eta) \le 0$ follows with $l = m+1$ and $x_{m+1} = x$.

Now, if ξ is optimal, $\max_\eta d(\xi, \eta) = 0$, and clearly $0 = d(\xi, \xi) = \sum_{i=1}^{m} \omega_i d(\xi, \xi_i)$ where $\xi_i = \{x_i; 1\}$, $i = 1, \ldots, m$. Hence $d(\xi, \xi_i) = 0$ for all $i = 1, \ldots, m$.

□

B.3 Proofs for Chapter 7

B.3.1 Theorem 8

Let $g \in \mathcal{G}_2$ and $\mathcal{X} = \{0, 1\}$. The minimax D-optimal two-point design supported at points 0 and 1 allocates a proportion ω^* of observations at point 0 regardless of the sign of the parameter β, where

$$\omega^* = \frac{\sqrt{\frac{c_2^2(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} + 1}\left[\sqrt{\frac{c_2^2(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} + 1} - \sqrt{\frac{c_2^2(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} + 1}\right]}{c_2^2\left[\frac{(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} - \frac{(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right]}.$$

Proof. If $g \in \mathcal{G}_2$ then $\left|\int_0^c e^{-y_j e^{\alpha+\beta x}} g(y_j)\, dy_j\right| \le c_2 \; \forall x \in \{0, 1\}$. Therefore, for a fixed design ξ supported at 0 and 1 with corresponding weights ω and $1-\omega$ the determinant of the mean squared error matrix defined in (7.10) is smaller than or equal to

$$\frac{1}{\omega(1-\omega)(1-e^{-ce^{\alpha}})(1-e^{-ce^{\alpha+\beta}})}\left\{1 + \omega\frac{(c_2 e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} + (1-\omega)\frac{(c_2 e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right\}$$

Taking the first order derivative of this expression with respect to ω and equating it to zero gives

$$\begin{aligned}
&\frac{(c_2 e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}\omega^2 - \frac{(c_2 e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}(1-\omega)^2 - (1-2\omega) = 0 \\
\iff \; &\omega_{1,2} = \frac{-\left[\frac{(c_2 e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} + 1\right] \pm \sqrt{\frac{(c_2 e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} + 1}\sqrt{\frac{(c_2 e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} + 1}}{c_2^2\left(\frac{(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} - \frac{(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right)}.
\end{aligned}$$

When β is positive, it is easy to check that both the numerator and the denominator of the above expression are non-positive since the function $\theta^2/(1-e^{-\theta})$ is increasing with θ. We reject the negative root of the numerator since

$$\begin{aligned}
&-\frac{(c_2 e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} - 1 - \sqrt{\frac{(c_2 e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} + 1}\sqrt{\frac{(c_2 e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} + 1} \\
&< -\frac{(c_2 e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} < c_2^2\left(\frac{(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} - \frac{(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right)
\end{aligned}$$

and the weight must always be always less than or equal to unity.

In the case of negative β-values the denominator is positive and since $\omega > 0$, again we accept the positive root.

Therefore for any sign of the parameter β the minimax D-optimal weight at point

0 is always given by

$$\omega^* = \frac{\sqrt{\frac{c_2^2(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}+1}\left[\sqrt{\frac{c_2^2(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}+1}-\sqrt{\frac{c_2^2(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}+1}\right]}{c_2^2\left[\frac{(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}-\frac{(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right]}.$$

□

B.3.2 Theorem 9(b)

Let $g \in \mathcal{G}_1$. If $\beta < 0$, the design with support points u and x_2^* and optimal weight on u given in (7.13) is minimax c-optimal for β on $\mathcal{X} = [u, v]$, where $x_2^* = v$ if

$$\beta(u-v) - \frac{2Q(\alpha+\beta u)}{Q'(\alpha+\beta u)}\left[1+\frac{\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta v)}}+\frac{4c_1^2 Q(\alpha+\beta u)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta v)}}\right] < 0.$$

Otherwise x_2^* is the unique solution of the equation

$$\beta(u-x_2) - \frac{2Q(\alpha+\beta x_2)}{Q'(\alpha+\beta x_2)}\left[1+\frac{\sqrt{Q(\alpha+\beta x_2)}}{\sqrt{Q(\alpha+\beta u)}}+\frac{4c_1^2 Q(\alpha+\beta x_2)\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta x_2)}}\right] = 0.$$

Proof. Let $\beta > 0$. Since $g \in \mathcal{G}_1$ then

$$\left|\int_0^c e^{-y_j e^{\alpha+\beta x}} g(y_j)\,dy_j\right| \le \int_0^c e^{-y_j e^{\alpha+\beta x}} |g(y_j)|\,dy_j \le \int_0^c e^{-y_j e^{\alpha+\beta x}} c_1\,dy_j$$
$$= c_1(1-e^{-ce^{\alpha+\beta x}})/e^{\alpha+\beta x}, \quad \forall x \in [u, v]$$

Therefore,

$$\left[\left|\frac{e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x1}} g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})} - \frac{e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x2}} g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}\right|\right]^2$$

$$\le \left[\left|\frac{e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x1}} g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}\right| + \left|\frac{e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x2}} g(y_j)\,dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}\right|\right]^2 \le 4c_1^2$$

Substituting the expression for the c-optimal weights from (7.13) we obtain the objective function to be minimised to be

$$\frac{1}{(x_1-x_2)^2}\left\{4c_1^2 + \left[\frac{1}{\sqrt{1-e^{-ce^{\alpha+\beta x_1}}}}+\frac{1}{\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}}\right]^2\right\} := k(x_1, x_2)$$

For fixed $x_2 (> x_1)$, $k(x_1, x_2)$ is increasing with x_1 as the product of two non-negative increasing functions and therefore attains its minimum at $x_1^* = u$. Now $k(u, x_2)$ has exactly one turning point on (u, ∞) which is a minimum since $\lim_{x_2 \to \infty} k(u, x_2) = \lim_{x_2 \to u} k(u, x_2) = \infty$.

If this minimum is attained outside $(u, v]$ then the larger support point of the design is v. This occurs if and only if

$$\left. \frac{\partial k(u, x_2)}{\partial x_2} \right|_{x_2 = v} < 0,$$

which is equivalent to condition

$$\beta(u - v) - \frac{2Q(\alpha + \beta u)}{Q'(\alpha + \beta u)} \left[1 + \frac{\sqrt{Q(\alpha + \beta u)}}{\sqrt{Q(\alpha + \beta v)}} + \frac{4c_1^2 Q(\alpha + \beta u)\sqrt{Q(\alpha + \beta v)}}{\sqrt{Q(\alpha + \beta u)} + \sqrt{Q(\alpha + \beta v)}} \right] < 0.$$

Otherwise, x_2^* is the larger support point and can be found by solving

$$\frac{\partial k(u, x_2)}{\partial x_2} = 0,$$

which is equivalent to solving equation

$$\beta(u - x_2) - \frac{2Q(\alpha + \beta x_2)}{Q'(\alpha + \beta x_2)} \left[1 + \frac{\sqrt{Q(\alpha + \beta x_2)}}{\sqrt{Q(\alpha + \beta u)}} + \frac{4c_1^2 Q(\alpha + \beta x_2)\sqrt{Q(\alpha + \beta u)}}{\sqrt{Q(\alpha + \beta u)} + \sqrt{Q(\alpha + \beta x_2)}} \right] = 0.$$

□

B.3.3 Theorem 10

Let $g \in \mathcal{G}_1$
(a) If $\beta > 0$, the design with support points x_1^* and v and optimal weight on x_1^* given in (7.17) is minimax c-optimal for α on $\mathcal{X} = [u, v]$, where $x_1^* = u$ if

$$\beta(u - v) + \frac{2Q(\alpha + \beta u)}{Q'(\alpha + \beta u)} \left[1 + \frac{\sqrt{Q(\alpha + \beta u)}}{\sqrt{Q(\alpha + \beta v)}} + \frac{2c_1^2 (u + v) Q(\alpha + \beta u)\sqrt{Q(\alpha + \beta v)}}{\sqrt{Q(\alpha + \beta u)} + \sqrt{Q(\alpha + \beta v)}} \right] > 0.$$

Otherwise x_1^* is the unique solution of the equation

$$\beta(x_1 - v) + \frac{2Q(\alpha + \beta x_1)}{Q'(\alpha + \beta x_1)} \left[1 + \frac{\sqrt{Q(\alpha + \beta x_1)}}{\sqrt{Q(\alpha + \beta v)}} + \frac{2c_1^2 (x_1 + v) Q(\alpha + \beta x_1)\sqrt{Q(\alpha + \beta v)}}{\sqrt{Q(\alpha + \beta x_1)} + \sqrt{Q(\alpha + \beta v)}} \right] = 0.$$

(b) If $\beta < 0$, the design with support points u and x_2^* and optimal weight on u given in (7.17) is minimax c-optimal for α on $\mathcal{X} = [u, v]$, where $x_2^* = v$ if

$$\beta(u-v) - \frac{2Q(\alpha+\beta u)}{Q'(\alpha+\beta u)}\left[1 + \frac{\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta v)}} + \frac{2c_1^2(u+v)Q(\alpha+\beta u)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta v)}}\right] < 0.$$

Otherwise x_2^* is the unique solution of the equation

$$\beta(u-x_2) - \frac{2Q(\alpha+\beta x_2)}{Q'(\alpha+\beta x_2)}\left[1 + \frac{\sqrt{Q(\alpha+\beta x_2)}}{\sqrt{Q(\alpha+\beta u)}} + \frac{2c_1^2(u+x_2)Q(\alpha+\beta x_2)\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta x_2)}}\right] = 0.$$

Proof. For $g \in \mathcal{G}_1$ we have that

$$\left|\int_0^c e^{-y_j e^{\alpha+\beta x}} g(y_j)\, dy_j\right| \leq \int_0^c e^{-y_j e^{\alpha+\beta x}} |g(y_j)|\, dy_j \leq \int_0^c e^{-y_j e^{\alpha+\beta x}} c_1\, dy_j$$
$$= c_1(1 - e^{-ce^{\alpha+\beta x}})/e^{\alpha+\beta x}, \qquad \forall x \in [u, v].$$

Therefore,

$$\left[\left|\frac{x_1 e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x_2}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})} - \frac{x_2 e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x_1}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}\right|\right]^2$$

$$\left[\left|\frac{x_1 e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x_2}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})} - \frac{x_2 e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x_1}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}\right|\right]^2$$
$$\leq \left[\left|\frac{x_1 e^{\alpha+\beta x_2}\int_0^c e^{-y_j e^{\alpha+\beta x_2}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_2}})}\right| + \left|\frac{x_2 e^{\alpha+\beta x_1}\int_0^c e^{-y_j e^{\alpha+\beta x_1}} g(y_j)\, dy_j}{(1-e^{-ce^{\alpha+\beta x_1}})}\right|\right]^2$$
$$\leq c_1^2(x_1+x_2)^2.$$

Substituting the expression for the c-optimal weights from (7.17) we obtain the objective function to be minimised to be

$$\frac{1}{(x_1-x_2)^2}\left\{c_1^2(x_1+x_2)^2 + \left[\frac{x_2}{\sqrt{1-e^{-ce^{\alpha+\beta x_1}}}} + \frac{x_1}{\sqrt{1-e^{-ce^{\alpha+\beta x_2}}}}\right]^2\right\} := \tilde{k}(x_1, x_2).$$

(a) Let $\beta > 0$. For fixed $x_1 (< x_2)$, $\tilde{k}(x_1, x_2)$ is decreasing with x_2 as the product of two non-negative decreasing functions and therefore attains its minimum at $x_2^* = v$. Now $\tilde{k}(x_1, v)$ has exactly one turning point on $(-\infty, v)$ which is a minimum since

$$\lim_{x_1 \to -\infty} \tilde{k}(x_1, v) = \lim_{x_1 \to v} \tilde{k}(x_1, v) = \infty.$$

If this minimum is attained outside $[u, v)$ then the smaller support point of the design is u. This occurs if and only if

$$\left.\frac{\partial \tilde{k}(x_1, v)}{\partial x_1}\right|_{x_1=u} > 0,$$

which is equivalent to condition

$$\beta(u-v) + \frac{2Q(\alpha+\beta u)}{Q'(\alpha+\beta u)}\left[1 + \frac{\sqrt{Q(\alpha+\beta u)}}{\sqrt{Q(\alpha+\beta v)}} + \frac{2c_1^2(u+v)Q(\alpha+\beta u)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta u)}+\sqrt{Q(\alpha+\beta v)}}\right] > 0.$$

Otherwise, x_2^* is the larger support point and can be found by solving

$$\frac{\partial \tilde{k}(u, x_2)}{\partial x_2} = 0,$$

which is equivalent to solving equation

$$\beta(x_1-v) + \frac{2Q(\alpha+\beta x_1)}{Q'(\alpha+\beta x_1)}\left[1 + \frac{\sqrt{Q(\alpha+\beta x_1)}}{\sqrt{Q(\alpha+\beta v)}} + \frac{2c_1^2(x_1+v)Q(\alpha+\beta x_1)\sqrt{Q(\alpha+\beta v)}}{\sqrt{Q(\alpha+\beta x_1)}+\sqrt{Q(\alpha+\beta v)}}\right] = 0.$$

The proof of part (b) follows along the same lines with similar arguments as for the proof of Theorem 9(b) and is therefore omitted.

□

References

Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, **10**, pp. 1100–1120.

Atkinson, A., Donev, A. and Tobias, R. (2007) *Optimum Experimental Designs, with SAS (Oxford Statistical Science Series).* Oxford University Press, USA, 2nd edn.

Bandyopadhyay, U., Biswas, A. and Bhattacharya, R. (2010) A covariate-adjusted adaptive design for two-stage clinical trials with survival data. *Statistica Neerlandica*, **64**, pp. 202–226.

Becker, N., McDonald, B. and Khoo, C. (1989) Optimal designs for fitting a proportional hazards regression model to data subject to censoring. *Australian Journal of Statistics*, **31**, pp. 449–468.

Biedermann, S., Dette, H. and Zhu, W. (2006) Optimal designs for doseresponse models with restricted design spaces. *Journal of the American Statistical Association*, **101**, pp. 747–759.

Biedermann, S. and Woods, D. C. (2011) Optimal designs for generalized non-linear models with application to second-harmonic generation experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **60**, pp. 281–299.

Breslow, N. (1974) Covariance analysis of censored survival data. *Biometrics*, **30**, pp. 89–99.

Chaloner, K. and Verdinelli, I. (1995) Bayesian experimental design: A review. *Statistical Science*, **10**, pp. 273–304.

Chernoff, H. (1953) Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, **24**, pp. 586–602.

Collett, D. (2003) *Modelling survival data in medical research.* London, New York: Chapman & Hall.

Cox, D. R. (1972) Regression models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, pp. 187–220.

— (1975) Partial likelihood. *Biometrika*, **62**, pp. 269–276.

Dette, H. (1997) Designing experiments with respect to 'standardized' optimality criteria. *Journal of the Royal Statistical Society. Series B (Methodological)*, **59**, pp. 97–110.

Dette, H. and Biedermann, S. (2003) Robust and efficient designs for the michaelis-menten model. *Journal of the American Statistical Association*, **98**, pp. 679–686.

Dette, H. and Melas, V. B. (2011) A note on the de la Garza phenomenon for locally optimal designs. *The Annals of Statistics*, **39**, pp 12661281.

Dette, H. and Sahm, K. (1998) Minimax optimal designs in nonlinear regression models. *Statistica Sinica*, **8**, pp 1249–1264.

Dette, H. and Schorning, K. (2013) Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *The Annals of Statistics*, **41**, pp 1260–1267.

Dror, H. A. and Steinberg, D. M. (2006) Robust experimental design for multivariate generalized linear models. *Technometrics*, **48**, pp. 520–529.

Efron, B. (1977) The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72**, pp. 557–565.

Elfving, G. (1952) Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics*, **23**, pp. 255–262.

— (1959) Design of linear experiments. In *Probability and Statistics. The Harald Cramér Volume* (ed. U. Grenander), pp. 58–74. Stockholm: Almagrist and Wiksell.

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. *Journal of the Royal Statistical Society. Series B (Methodological)*, **54**, pp. 569–583.

Freireich, E. J., Gehan, E., Frei III, E., Schroeder, L. R., Wolman, I. J., Anbari, R., Burgert, O. E., Mills, S. D., Pinkel, D., Selawry, O. S., Moon, J. H., Gendel, B. R., Spurr, C. L., Storrs, R., Haurani, F., Hoogstraten, B. and Lee, S. (1963) The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukimia: A model for evaluation of other potentially useful therapy. *Blood*, **21**, pp. 699–716.

Haines, L. M. (1995) A geometric approach to optimal design for one-parameter non-linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, pp. 575–598.

Hedayat, A., Zhong, J. and Nie, L. (2004) Optimal and efficient designs for 2-parameter nonlinear models. *Journal of Statistical Planning and Inference*, **124**, pp. 205–217.

Imhof, L. A. (2001) Maximin designs for exponential growth models and heteroscedastic polynomial models. *The Annals of Statistics*, **29**, pp. 561–576.

Imhof, L. A. and Wong, W. K. (2000) A graphical method for finding maximin efficiency designs. *Biometrics*, **56**, pp. 113–117.

Kalbfleisch, J. D. and Prentice, R. L. (2002) *The Statistical Analysis of Failure Time Data*. New York: Wiley-Interscience, 2nd edn.

Kalish, L. A. and Harrington, D. P. (1988) Efficiency of balanced treatment allocation for survival analysis. *Biometrics*, **44**, pp. 815–821.

Kiefer, J. (1974) General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, **2**, pp. 849–879.

Konstantinou, M., Biedermann, S. and Kimber, A. C. (2013) Optimal designs for two-parameter nonlinear mdoels with application to survival models. *Statistica Sinica, in press*. doi:10.5705/ss.2011.271.

López-Fidalgo, J. and Rivas-López, M. J. (2012) Optimal experimental designs for partial likelihood information. *Computational Statistics & Data Analysis*. doi:10.1016/j.csda.2012.10.009.

López-Fidalgo, J., Rivas-López, M. J. and Del Campo, R. (2009) Optimal designs for cox regression. *Statistica Neerlandica*, **63**, pp. 135–148.

McGree, J. and Eccleston, J. (2010) Investigating design for survival models. *Metrika*, **72**, pp. 295–311.

Pascual, F. G. and Montepiedra, G. (2003) Model-robust test plans with applications in accelerated life testing. *Technometrics*, **45**, pp. 47–57.

Prékopa, A. (1973) On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, **33**, 335–343.

Pukelsheim, F. and Rieder, S. (1992) Efficient rounding of approximate designs. *Biometrika*, **79**, pp. 763–770.

Pukelsheim, F. and Torsney, B. (1991) Optimal weights for experimental designs on linearly independent support points. *The Annals of Statistics*, **19**, pp. 1614–1625.

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

Sebastiani, P. and Settimi, R. (1997) A note on D-optimal designs for a logistic regression model. *Journal of Statistical Planning and Inference*, **59**, pp. 359–368.

Sibson, R. (1972) Discussion on results in the theory and construction of D-optium experimental designs (by H. P. Wynn). *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, pp. 174–175.

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

Sinha, S. and Wiens, D. P. (2002) Robust sequentlial designs for nonlinear regression. *The Canadian Journal of Statistics*, **30**, pp. 601–618.

Sitter, R. R. and Torsney, B. (1995a) *D*-optimal designs for generalized linear models. In *MODA 4 - Advances in model-oriented data analysis* (eds. C. P. Kitsos and W. G. Muller), 87–102. Heidelberg: Physica-Verlag.

— (1995b) Optimal designs for binary response experiments with two design variables. *Statistica Sinica*, **5**, pp. 405–419.

Wiens, D. P. (1992) Minimax designs for approximately linear regression. *Journal of Statistical Planning and Inference*, **31**, pp. 353 – 371.

Wiens, D. P. and Xu, X. (2008a) Robust designs for one-point extrapolation. *Journal of Statistical Planning and Inference*, **138**, pp. 1339 – 1357.

— (2008b) Robust prediction and extrapolation designs for misspecified generalized linear regression models. *Journal of Statistical Planning and Inference*, **138**, pp. 30–46.

Woods, D. C., Lewis, S. M., Eccleston, J. A. and Russell, K. G. (2006) Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, **48**, pp. 284–292.

Wu, S. J., Lin, Y. P. and Chen, Y. J. (2006) Planning step-stress life test with progressively type i group-censored exponential data. *Statistica Neerlandica*, **60**, pp. 46–56.

Xu, X. (2009a) Robust designs for misspecified exponential regression models. *Applied Stochastic Models in Business and Industry*, **25**, pp. 179–193.

— (2009b) Robust prediction and extrapolation designs for censored data. *Journal of Statistical Planning and Inference*, **139**, pp. 486–502.

Yang, M. (2010) On the de la Garza phenomenon. *The Annals of Statistics*, **38**, pp 2499–2524.

Yang, M. and Stufken, J. (2009) Support points of locally optimal designs for nonlinear models with two parameters. *The Annals of Statistics*, **37**, pp. 518–541.

— (2012) Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *The Annals of Statistics*, **40**, pp 1665–1681.

Zhang, Y. and Meeker, W. Q. (2006) Bayesian methods for planning accelerated life tests. *Technometrics*, **48**, pp. 49–60.