Institut für Mathematik
Mathematische Statistik

Modifications and Extensions of the Logistic Regression and Cox Model

Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin "Mathematische Statistik"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von
Korakot Wichitsa-nguan

Potsdam, den 04.01.2016

Acknowledgements

This thesis would not have been possible without the help and support of many people.

Foremost, I would like to express my deepest gratitude to my supervisor Prof. Dr. Hannelore Liero for the continuous support of my Ph.D study and research, for encouragement, great patience, motivation, enthusiasm, and immense knowledge. Without her guidance and persistent help this thesis would not have been possible. One simply could not wish for a better or friendlier supervisor.

I would also like to thank Prof.Dr.Henning Läuter for his timely suggestions with kindness, comments and advice throughout my thesis with his patience and knowledge. Especially, when I first studied and had trouble understanding the previous work.

In addition, I am grateful to Prince of Songkla University for giving me the financial support for the first three years and the opportunity to study in Germany. Particularly, I am thankful to the Department of Mathematics and Statistics, Prince of Songkla University for giving a great knowledge in mathematics, teaching experience and patiently waiting for me.

I am indebted to all of my colleagues and ex-colleagues for their suggestions about studying and living in Germany. Moreover, I would like to give very special thanks to my dear friend, Sutthikarn Kamales and Treewut Rassamegevanon for their advice in proofreading.

Thanks are also extended to all of my friends in Germany for helping me, supporting me, and especially sharing the experiences throughout my journey away from home. I will always appreciate the encouragement and precious moments we shared.

Last but not the least, special recognition goes out to my parents Mr.Siriroj and Ms.Kanlaya, for all their love, constant encouragement, and always supporting me spiritually throughout my life. My work would not have been accomplished without their encouragement and understanding.

Contents

Acknowledgements **i**

List of Symbols **ix**

1 Introduction **1**

2 Logistic Regression under Case-Cohort Design **9**

2.1 Introduction 9

2.2 Logistic regression with complete data: A short summary 10

2.2.1 Logistic regression model 10

2.2.2 Logistic regression and 2×2 tables 12

2.3 Estimators in case-cohort models with discrete covariates . . 15

2.3.1 The maximum likelihood estimator in 2×2-tables and their properties 16

2.3.2 The maximum likelihood estimator in logistic regression with discrete covariates 24

3 Cox Models and their extensions: Survey of approaches and results **33**

3.1 Introduction 33

3.2 The Cox model (Proportional hazards model) 34

3.2.1 Partial likelihood estimator and its properties 35

3.3 The Cox model with time-dependent covariates 38

3.3.1 Counting process approach to the Cox model 41

3.4 The Cox model with time-dependent coefficients 47

4 Estimability of the parameter in the Cox model and optimal choice of the covariates 51
4.1 Estimability of the parameter $\boldsymbol{\beta}_0$ 52
4.1.1 Examples 58
4.1.2 Asymptotic estimability 61
4.1.3 Asymptotic variance for general covariates 66
4.2 Optimal covariates 69
4.2.1 Local optimal covariates 70
4.2.2 Sensitivity analysis 77
4.3 Proposals for applicable optimal covariates 80
4.3.1 Two-stage choice of optimal covariates 80
4.3.2 Nearly D-optimal covariates 81
4.3.3 Mixed optimal covariates 82
4.4 Suboptimal covariates 82
4.4.1 Iterative calculation of suboptimal covariates 84

5 Statistical inference in the Cox model with time-dependent coefficient 91
5.1 The maximum local partial likelihood estimation 92
5.1.1 Simulation Study 95
5.2 Score test based on the local partial likelihood approach . . . 97
5.2.1 Distribution of the quadratic form of the score vector . 98
5.2.2 Formulation of the test procedure 106
5.3 Bootstrap version of the score test 110
5.3.1 Simulation Study 112

A Figures and Tables for Chapter 4 115

List of Figures

2.1 Conceptual illustration of the immunogenome and cancer case-cohort design. The proportional of numbers of observations are represented by areas of different colors. 16
2.2 Histogram of β^* estimates in full cohort. 22
2.3 Histogram of β^* estimates under the case-cohort design. 23
2.4 The plot of function $\pi(x)$ under $\beta_0 = -3$ and $\beta_1 = 1.5$ with several x. 27

5.1 Estimation of $\beta_1(t) = t$ and $\beta_2(t) = \frac{1}{2}$ for Model I with $n = 600$, $h = 0.6$ and 30% censoring. The red lines are the true functions. 96
5.2 Estimation of $\beta_1(t) = t$ and $\beta_2(t) = \frac{1}{2}$ for Model I with $n = 600$, $h = 0.8$ and 30% censoring. The red lines are the true functions. 97
5.3 Estimation of $\beta(t) = \log(t)$ for Model II with $n = 800$, $h = 0.75$ and 30% censoring. The blue lines are the true functions. 98
5.4 The score function at 2 grid points, 200 simulations 104
5.5 The estimated cumulative baseline hazard function and the estimated baseline survival function under $\mathcal{H}$. 111
5.6 Empirical distribution of p-value for the score test in the first simulation, i.e. $\beta_0 = 2$ (left plot) and the second simulation, i.e. $\beta_0(t) = \log(t)$ (right plot) 114

A.1 The inverse of the asymptotic variance for the exponential model in dependence of the value of ξ_2 for several ξ_1. The values for $\lambda_0 = 1$ and $q_1 = 0.5$ are fixed, β_0 takes three different values. 116

A.2 The inverse of the asymptotic variance for the exponential model in dependence of the value of ξ_2 for several ξ_1. The values for $\lambda_0 = 3$ and $q_1 = 0.5$ are fixed, β_0 takes three different values. 117
A.3 The inverse of the asymptotic variance with $\beta_0 = 1.0$ (the left hand side) and $\beta_0 = 6.0$ (the right hand side) for the exponential model, $\lambda_0 = 1$, in pair (ξ_1, ξ_2) where $\xi_1 = 0.8, 0.9, 1$ and $\xi_2 = 0, 0.1, 0.2, 0.3, 0.4$ with the different q_1. 118
A.4 The inverse of the asymptotic variance with β_0 for the Weibull model, $\theta_0 = 1, \mu = 2$, $q_1 = 0.5$ in every ξ_1 with the different ξ_2 119
A.5 The inverse of the asymptotic variance with β_0 for the Weibull model, $\theta_0 = 1, \mu = 5$, $q_1 = 0.5$ in every ξ_1 with the different ξ_2 120
A.6 The inverse of the asymptotic variance with β_0 for the Weibull model, $\theta_0 = 3, \mu = 2$, $q_1 = 0.5$, in every ξ_1 with the different ξ_2. 121
A.7 The inverse of the asymptotic variance with β_0 for the Weibull model, $\theta_0 = 3, \mu = 5$, $q_1 = 0.5$, in every ξ_1 with the different ξ_2. 122

List of Tables

2.1 Numbers of individuals in the immunogenome study cohort and case-cohort. 16
2.2 Simulation summary statistics for estimation under case-cohort design with different values of α. 22
2.3 The variance estimator of case cohort based on α. 24
2.4 The estimates results of logistic regression with 3 different values of covariate. 27
2.5 The theoretical values and the simulation values of the asymptotic variance with different values of α, f_1 and f_2. 32

4.1 Local optimal two-points covariates of the exponential model (4.17) as a submodel of (3.2). 73
4.2 Local optimal two-points covariates of the Weibull model (4.23) as a submodel of (3.2). 77
4.3 Local suboptimal covariates for the exponential model as a submodel of (3.2) 88

A.1 Local optimal covariates for $\tilde{\beta}_0 = 1$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \delta_{[0,\tau)}(t)$ for $t \geq 0$. 123
A.2 Local optimal covariates for $\tilde{\beta}_0 = 1$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \exp(-t)$. 123
A.3 Local optimal covariates for $\tilde{\beta}_0 = 1$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = (\exp(ct) - 1)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}\log 2$. 124
A.4 Local optimal covariates for $\tilde{\beta}_0 = 1$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = (ct)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}$. 124
A.5 Local optimal covariates for $\tilde{\beta}_0 = 2$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \delta_{[0,\tau)}(t)$ for $t \geq 0$. 125

A.6 Local optimal covariates for $\tilde{\beta}_0 = 3$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \delta_{[0,\tau)}(t)$ for $t \geq 0$. 125
A.7 Local optimal covariates for $\tilde{\beta}_0 = 3$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \exp(-t)$. 126
A.8 Local optimal covariates for $\tilde{\beta}_0 = 3$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = (\exp(ct) - 1)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}\log 2$. 126
A.9 Local optimal covariates for $\tilde{\beta}_0 = 3$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = (ct)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}$. 127

List of Symbols

The following notation is used throughout the thesis:

$\mathbb{R}$	real number
$\mathbb{R}^m$	m-dimensional vectors of real number
$\mathfrak{B}^m$	Borel-σ algebra over $\mathbb{R}^m$
$\lvert\boldsymbol{x}\rvert$	Euclidean norm of vector $\boldsymbol{x}$
$\lVert\lambda\rVert$	sup norm of function λ
$\mathbb{1}(A)$	indicator function of the set A
P	probability measure
$\mathsf{E}(X)$	expectation of a real random variable X
$\mathsf{Var}\ (X)$	variance of a real random variable X
$\mathsf{N}(m,\sigma^2)$	normal distribution with expectation m and variance σ^2
$\mathsf{U}(a,b)$	uniform distribution on the interval [a,b]
$\mathsf{Exp}(a)$	exponential distribution with rate a
χ^2_d	Chi-square distribution with degree of freedom d
rank(B)	rank of matrix B
$C([0,\tau])$	continuous functions over interval $[0,\tau]$
$\xrightarrow{\mathrm{P}}$	convergence in probability
$\xrightarrow{\mathrm{D}}$	convergence in distribution (weak convergence of probability measure)
$\mathcal{R}(t)$	risk set at time t
K	kernel function, $K_a(x) = (1/a)K(\frac{x}{a})$, $\kappa^2 = \int K^2(x)dx$
$\mathcal{H}$	null hypothesis
$\mathcal{K}$	alternative hypothesis

Chapter 1

Introduction

In many statistical applications, the aim is to model the relationship between covariates and some outcomes. A choice of the appropriate model depends on the outcome and the research objectives, such as linear model for continuous outcomes, logistic models for binary outcomes and the Cox model for time-to-event data.
In epidemiological, medical, biological, societal and economic studies, the logistic regression is widely used to describe the relationship between a response variable as binary outcome and explanatory variables as a set of covariates.
In particular, a binary response variable, such as “disease or non-disease”, “success or failure”, “enrolled or not enrolled”, “presence or absence”, is mostly considered. For example, Arnlöv et al. (2010) considered modelling the presence or absence of metabolic syndrome using the impact of body mass index (BMI) as a explanatory variable.
The effect of the covariate on the occurrence of the event of interest is described by a model for the conditional probability of the outcome given the covariates. That is, the response variable, say Y, is the indicator of the event of interest and the task is to model the function

$$\pi(\boldsymbol{x}) = \mathsf{P}(Y = 1|\boldsymbol{X} = \boldsymbol{x}) = \mathsf{E}(Y|\boldsymbol{X} = \boldsymbol{x})$$

where $\boldsymbol{X}$ is a p-dimensional explanatory variable.
A widely used model is the logistic regression. Define the odds on covariates $\boldsymbol{x}$ as

$$\frac{\mathsf{P}(Y = 1|\boldsymbol{X} = \boldsymbol{x})}{\mathsf{P}(Y = 0|\boldsymbol{X} = \boldsymbol{x})} = \frac{\pi(\boldsymbol{x})}{1 - \pi(\boldsymbol{x})}.$$

In the logistic regression model the logarithm of the odds, so-called logit, is modeled by a linear relation:

$$\log\left(\frac{\pi(\boldsymbol{x},\boldsymbol{\beta})}{1-\pi(\boldsymbol{x},\boldsymbol{\beta})}\right) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j = \boldsymbol{x}^T\boldsymbol{\beta}.$$

The coefficients β_j can be interpreted as follows:
The odds ratio of a variable X_j adjusted for the other variables X_k, $j,k = 1,\ldots,p$, $j \neq k$ is given by

$$\frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_j(x_j+1) + \ldots + \beta_p x_p)}{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j + \ldots + \beta_p x_p)} = \exp(\beta_j). \tag{1.1}$$

In other words, $\exp(\beta_j)$ characterizes the factor by which the odds of the event changes for each one unit increase of X_j (similar to the simple linear regression).
The goal of logistic regression is to estimate the parameter $\boldsymbol{\beta}$ (like in ordinary regression). The estimation in logistic regression makes use of the likelihood function.
However, epidemiologic cohort studies are quite expensive to manage data because we have to follow up a large number of individuals for a long time.
The case-cohort design is applied to reduce the cost and time for data collection. The case-cohort sampling collects a small random sample from the entire cohort, is called *subcohort*. The advantage of this design is that the covariate and follow-up data are recorded only on the subcohort and all cases.
The aim in the present thesis is to develop an estimation approach in the logistic model to case-cohort design.

The first part of this thesis is presented as follows:

Chapter 2: *Logistic Regression under Case-Cohort Design.*
This chapter investigates the estimation in the logistic model for case-cohort design. First a model with a binary response and a binary covariate X (which is also called exposure variable) under case-cohort design is considered.
We describe the maximum likelihood estimator (MLE) and establish the consistency and the asymptotic normality of this estimator.

An estimator for the asymptotic variance of the estimator based on the maximum likelihood approach is proposed; this estimator differs slightly from the estimator introduced by Prentice (1986). Simulation results for several proportions of the subcohort show that the proposed estimator gives lower empirical bias and empirical variance than Prentice's estimator.
The MLE in the logistic regression with discrete covariate under case-cohort design is studied. Here the approach of the binary covariate model is extended. Proving asymptotic normality of estimators, standard errors for the estimators can be derived.
The estimation procedure of the logistic regression model with a one-dimensional discrete covariate is demonstrated by simulation. The results of the simulation for several proportions of the subcohort and different choices of the underlying parameters illustrate that the estimate values are around the true values.
Moreover, the comparison between theoretical values and simulation results of the asymptotic variance of estimator is presented in the last part of this chapter.

Epidemiology and medical studies are often interested in time to occurrence of an event. Logistic regression is sufficient for the binary outcome refers to be available for all subjects and for a fixed time interval.
In practice, the observations in clinical trials are frequently collected for different time periods and subjects may drop out or relapse from other causes during follow-up. The survival analysis is necessary to solve these problems. In survival analysis, the subjects are followed over time and the time to event of interest $\tilde{T}$ is focused on these studies as response variable in binary regression Y.
The characterization of the distribution of $\tilde{T}$ is given by the hazard function, defined as

$$\lambda(t|\boldsymbol{x}) = \lim_{\triangle t \downarrow 0} \frac{1}{\triangle t} \mathsf{P}(t \leq \tilde{T} < t + \triangle t | \tilde{T} \geq t, \boldsymbol{X} = \boldsymbol{x}).$$

Cox (1972) proposed the model which is focused on the hazard function. The Cox model is assumed to be

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x})$$

where $\lambda_0(t)$ is an unspecified baseline hazard at time t and $\boldsymbol{X}$ is covariates, $\boldsymbol{\beta}$ is a p-dimensional vector of coefficient.
The hazard ratio for any two sets of covariates $\boldsymbol{x}$ and $\boldsymbol{x}^*$ is

$$\frac{\lambda(t|\boldsymbol{x})}{\lambda(t|\boldsymbol{x}^*)} = \frac{\lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{x})}{\lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{x}^*)} = \exp(\boldsymbol{\beta}^T(\boldsymbol{x}-\boldsymbol{x}^*)), \text{ for all } t \geq 0$$

which is a constant over time, i.e. the hazards are proportional to each other.
If X_j adjusted for the other covariates X_k, $j,k = 1,\ldots,p$, $j \neq k$, the hazard ratio for each one unit increase of covariate is given by

$$\frac{\lambda(t|x_1,\ldots,(x_j+1),\ldots,x_p)}{\lambda(t|x_1,\ldots,x_j,\ldots,x_p)} = \exp(\beta_j), \text{ for all } t \geq 0.$$

it is similar to (1.1). Therefore, the coefficient β_j is the log hazard ratio between two subject differing by one unit.
While the logistic regression estimates the odds ratio, Cox regression estimates the hazard ratio.
However, the logistic regression is not appropriate for incomplete follow-up data; for example, an individual drops out of the study before the end of data collection or an individual has not occurred the event of interest in the duration of the study. These observations are called censored observations. The Cox models can effectively handle censored data. Moreover, the Cox models can be used for discrete or continuous values of covariates. This leads to the second part where the Cox model is developed as a problem of experimental design in Chapter 4. The Cox models and their extensions are first described in Chapter 3. Particularly, the extended Cox model with time-dependent covariates and time-dependent coefficients are introduced. And then the statistical inference in Cox model with time-dependent coefficients are proposed in Chapter 5.
Three chapters in the second part are presented as follows:

Chapter 3: *Cox Models and their extensions: Survey of approaches and results.*
This chapter gives a review of the literature which includes a description of survival analysis framework that has been used to provide appropriate modifications to statistical inference procedure used in Cox model.
We start by describing the basic notion of survival analysis and introducing the Cox model. The maximum partial likelihood estimator (MPLE) and its

consistency and asymptotic normality are investigated. Test statistics for testing hypotheses about the true parameter value of $\boldsymbol{\beta}$, say $\boldsymbol{\beta}_0$, in the Cox model are also presented.
Furthermore, we introduce the extension of the Cox model in which the covariates $\boldsymbol{X}$ are allowed to depend on time, such covariates are called *time-dependent covariates* or *time-varying covariates*. Another popular extension of the Cox model is so-called *time-dependent coefficients* Cox model where the coefficients $\boldsymbol{\beta}$ are extended to vary with time. In general, these coefficients often depend on time and need to be tested. This leads to Chapter 5 where the statistical inference in the Cox model is developed in order to account for time-dependent coefficients.
In this chapter, we also provide the counting process framework to establish the asymptotic properties of the estimator of two different extensions of Cox model. The observed information matrix $\mathbf{I}_n(\boldsymbol{\beta})$ and asymptotic variance matrix $\boldsymbol{\Sigma}^{-1}$ are also obtained.

Chapter 4: *Estimability of the parameter in the Cox model and optimal choice of the covariates.*
In this chapter the estimability of the parameter $\boldsymbol{\beta}_0$ in the Cox model, where $\boldsymbol{\beta}_0$ denotes the true value of $\boldsymbol{\beta}$, and the choice of optimal covariates are investigated. We give new representations of the observed information matrix $\mathbf{I}_n(\boldsymbol{\beta})$ and extend results for the Cox model of Andersen and Gill (1982). In this way, conditions for the estimability of $\boldsymbol{\beta}_0$ are formulated. Here the results about the rank of $\mathbf{I}_n(\boldsymbol{\beta})$ are important, that is, if $\mathbf{I}_n(\boldsymbol{\beta})$ has the full rank p, then the parameter $\boldsymbol{\beta}_0$ is estimable. Moreover, we say that $\boldsymbol{\beta}_0$ is asymptotically estimable, if the limiting matrix of $\mathbf{I}_n(\boldsymbol{\beta})$, more exactly $\boldsymbol{\Sigma} = \operatorname{plim}_{n\to\infty} n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0)$, has the full rank p. With these results on the estimability of $\boldsymbol{\beta}_0$ we find the connection to estimation problems in other classical statistical models.
Under some regularity conditions $\boldsymbol{\Sigma}$ is the inverse of the asymptotic variance matrix of the MPLE of $\boldsymbol{\beta}_0$ in the Cox model. This $\boldsymbol{\Sigma}$ is the basis for other statistical analyses. In Theorem 4.4 a representation of $\boldsymbol{\Sigma}$ is found. The explicit dependence of $\boldsymbol{\Sigma}$ on $\boldsymbol{\beta}_0$, λ_0, G and Q is visible in which G is the censoring distribution and Q is the distribution of the covariates. We describe that for the Cox model the asymptotic estimability depends only on the support points of the covariates $\boldsymbol{X}$ where $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m$ are the different support points of the covariates. We show in Theorem 4.6 that the matrix

$\mathbf{\Sigma}$ is non-singular if and only if such points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m$ in the support of Q exist that $\{\boldsymbol{\xi}_r - \boldsymbol{\xi}_s, 1 \leq s < r \leq m\}$ spans the $\mathbb{R}^p$. Some representations of $\mathbf{\Sigma}$ are proved under the assumption that the support of the covariates is finite, but we find also the representation of $\mathbf{\Sigma}$ for a general distribution Q. Moreover, some properties of the asymptotic variance matrix of the MPLE are highlighted.

The explicit dependence of $\mathbf{\Sigma}$ on the distribution Q derived as above is necessary for finding optimal covariates. Optimal covariates under a survival framework were considered in López-Fidalgo et al. (2009), Garcet-Rodr´i guez et al. (2008), Balakrishnan and Han (2007) and Schmidt and Schwabe (2015). In these papers the authors used the maximum likelihood estimator (MLE), but in the Cox model the MPLE should be the basis for statistical analyses. Our received new representation of the asymptotic variance matrix $\mathbf{\Sigma}^{-1}$ or of the $\mathbf{\Sigma}$ gives the possibility for characterizing optimal covariates.

In our approach for finding optimal covariates for the MPLE, we have similar problems as in the experimental design for least squares estimate in nonlinear regression. We see that $\mathbf{\Sigma}$ and consequently the asymptotic variance matrix depend on the unknown parameters of the model. Hence, local optimal covariates are investigated and also calculated in examples. We remark that the optimal covariates depend strongly on the parameter $\boldsymbol{\beta}_0$, but weakly on the baseline hazard function λ_0.

In a sensitivity analysis the efficiency of given covariates is calculated. We find for neighborhoods of the exponential models the efficiencies and see that for fixed parameters $\boldsymbol{\beta}_0$ the efficiencies do not change very much for different baseline hazard functions.

In section 4.3 some proposals for applicable optimal covariates are discussed. We consider two-stage optimal covariates where the sample is divided in two parts. The second part is taken as optimal covariates if one uses the estimator calculated with the observations from the first part. Furthermore, we consider covariates where a grid in the parameter space is fixed and a weighted sum of a function of the corresponding $\mathbf{\Sigma}$-values is to be maximized.

Finally, a calculation procedure for finding optimal covariates is discussed. Generally, one obtains suboptimal covariates.

In this chapter, several results about optimal covariates are obtained. In particular, the results about the $\mathbf{I}_n(\boldsymbol{\beta})$, $\mathbf{\Sigma}$ and the corresponding proved properties extend the up to now known results in many ways.

Chapter 5: *Statistical inference in the Cox model with time-dependent coefficients.*
In this chapter we focus on the Cox model with time-dependent coefficients:

$$\lambda(t, \boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T(t)\boldsymbol{x}),$$

where $\boldsymbol{\beta}(\cdot)$ is a p-dimensional vector of time-dependent coefficients function. The maximum local partial likelihood estimators for estimating the coefficients function $\boldsymbol{\beta}(\cdot)$ are described.
The maximum local likelihood method for the nonparametric estimation in survival regression was introduced by Tibshirani and Hastie (1987). The idea is to estimate the coefficient function at a grid point t by maximizing the log partial likelihood function locally in a window around t. The local neighborhood is described by a kernel function, and the size of the neighborhood is controlled by a smoothing parameter h.
First we give two examples for realization of such estimates. The results show that the both estimates are quite close to the true functions for different choices of bandwidth and different censoring patterns.
The main topic of this chapter is a new test procedure for testing whether a one-dimensional coefficient function $\beta(\cdot)$ has a prespecified parametric form, say $\beta(\cdot, \vartheta)$. The test procedure is derived as follows: We consider the score function derived from the local constant partial likelihood function at d distinct grid points. It is shown that the distribution of the properly standardized quadratic form of this d-dimensional vector at the hypothetical $\beta(\cdot; \vartheta_0)$ tends in distribution to a Chi-squared distribution.
Moreover, replacing the unknown ϑ_0 by the MPLE in the hypothetical model the limit statement remains true, and an asymptotic α-test is given by the quantiles or p-values of the limiting Chi-squared distribution.
It is known that the performance of an asymptotic α-test using estimates based on smoothing depends not only on the sample size n but also on a suitable choice of the smoothing parameter, here h.
So it seems to be useful to give also a bootstrap version of this test. To derive such a bootstrap procedure, the resampling method developed by Davison and Hinkley (1997) for bootstrapping survival data is applied. The bootstrap test is only defined for the special case of testing whether the coefficient function is constant.
A simulation study which illustrates the behavior of the bootstrap test under the null hypothesis and a special alternative gives good results for the chosen underlying model.

Chapter 2

Logistic Regression under Case-Cohort Design

2.1 Introduction

Regression models describe the relationship between a response variable and one or more explanatory variables. If the outcome is a binary variable, logistic models are often used. These models belong to the generalized linear models (GLM). By the logit function, the conditional expectation of Y is transformed to a linear function, in other words we have for the conditional probability

$$\pi(\boldsymbol{x}, \boldsymbol{\beta}) = \mathsf{P}(Y = 1|\boldsymbol{X} = \boldsymbol{x}) = \mathsf{E}(Y|\boldsymbol{X} = \boldsymbol{x})$$

$$\log \frac{\pi(\boldsymbol{x}, \boldsymbol{\beta})}{1 - \pi(\boldsymbol{x}, \boldsymbol{\beta})} = \beta_0 + \sum_{j=1}^{p} \beta_j x_j = \boldsymbol{x}^T \boldsymbol{\beta} \tag{2.1}$$

or equivalently

$$\pi(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}^T \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}^T \boldsymbol{\beta})}$$

with $\boldsymbol{x} = (1, x_1, \ldots, x_p)^T$.

Often $\boldsymbol{X}$ is a so-called exposure variable, and one defines the odds on exposure $\boldsymbol{x}$ as

$$\psi(\boldsymbol{x}) = \frac{\pi(\boldsymbol{x}, \boldsymbol{\beta})}{1 - \pi(\boldsymbol{x}, \boldsymbol{\beta})}$$

and

$$\log \psi(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\beta}$$

are the log odds. For $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ with $x_{2j} = x_{1j}$ for $j = 1, \ldots, p$, $j \neq k$ the odds ratio and the log odds ratio is given by

$$\frac{\psi(\boldsymbol{x}_1)}{\psi(\boldsymbol{x}_2)} = \exp((x_{1k} - x_{2k})\beta_k) \qquad \log\frac{\psi(\boldsymbol{x}_1)}{\psi(\boldsymbol{x}_2)} = (x_{1k} - x_{2k})\beta_k,$$

respectively. Thus, for $x_{1k} = 0$ and $x_{2k} = 1$ we get

$$\beta_k = \log\frac{\psi(\boldsymbol{x}_1)}{\psi(\boldsymbol{x}_2)}.$$

We will consider a case-cohort sampling scheme. In a case-cohort study one selects at the beginning of the study a subcohort which is a simple sample of the entire cohort. The covariates are measured only for this random subcohort and for all cases–cases are all members of the cohort developing the event of interest during the follow-up. This sampling design allows the estimation of the odds and of the conditional probabilities $\pi(\boldsymbol{x})$.
Notice that this sampling scheme differs from case-control studies. A case-control study takes sampling separately from cases (individuals who developed the disease) and controls (individuals without disease). The previous exposure of both cases and controls are determined. Here odds ratios can be estimated, however the probability $\pi(\boldsymbol{x})$ cannot be estimated.

2.2 Logistic regression with complete data: A short summary

2.2.1 Logistic regression model

Suppose that we observe i.i.d. random variables $(Y_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$. Usually in logistic models the distribution of the covariates is not specified. The likelihood approach is based on the conditional distribution of the response, given $\boldsymbol{X} = \boldsymbol{x}$. Since $Y|\boldsymbol{X} = \boldsymbol{x}$ is distributed according to Bernoulli distribution with parameter $\pi(\boldsymbol{x})$ the likelihood function has the following form:

$$L(\boldsymbol{\beta}) \;=\; \prod_{i=1}^{n} \left(\pi(\boldsymbol{x}_i, \boldsymbol{\beta})^{Y_i} \left(1 - \pi(\boldsymbol{x}_i, \boldsymbol{\beta})\right)^{1-Y_i}\right),$$

and the log likelihood function by

$$\begin{aligned}\ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \log\left(\pi(\boldsymbol{x}_i, \boldsymbol{\beta})^{Y_i} \, (1 - \pi(\boldsymbol{x}_i, \boldsymbol{\beta}))^{1-Y_i}\right) \\ &= \sum_{i=1}^{n} \sum_{j=0}^{p} Y_i \, x_{ij} \beta_j \; - \; \sum_{i=1}^{n} \log(1 + \exp(\sum_{j=0}^{p} x_{ij} \beta_j)).\end{aligned}$$

The estimator $\widehat{\boldsymbol{\beta}}_n$ is the solution of the score equations

$$u_j(\boldsymbol{\beta}) = 0 \qquad j = 0, \ldots, p \tag{2.2}$$

with

$$u_j(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} \; = \; \sum_{i=1}^{n} x_{ij} \left(Y_i - \pi(\boldsymbol{x}_i, \boldsymbol{\beta})\right).$$

Note that in general (2.2) has no explicit solution and a numeric procedure is necessary to determine the value of $\widehat{\boldsymbol{\beta}}_n$.
The asymptotic properties of this estimator follow from the general theory of asymptotic properties of maximum likelihood estimators in GLM. Under weak regularity conditions $\widehat{\boldsymbol{\beta}}_n$ is consistent and asymptotically normal. Based on this result, the Wald test, the likelihood-ratio test and the score test for testing hypotheses of the coefficients β_j can be derived. For later reference we will give here the conditional Fisher information matrix. This matrix can be used for the iterative computation of the estimator; moreover, replacing the unknown $\boldsymbol{\beta}$ in $\mathcal{I}_n$, estimates for the standard errors of the coefficients are derived by inverting the estimated Fisher information.

The elements of the conditional observed information matrix are equal to the conditional information matrix; they are given by

$$\begin{aligned}\mathcal{I}_{n\,rs}(\boldsymbol{\beta}) &= -\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_r \, \partial \beta_s} \qquad\qquad r, s = 0, \ldots, p \\ &= \sum_{i=1}^{n} \pi(\boldsymbol{x}_i, \boldsymbol{\beta})(1 - \pi(\boldsymbol{x}_i, \boldsymbol{\beta})) x_{ir} x_{is}.\end{aligned}$$

Under the assumption that the matrix $n^{-1}\mathcal{I}_n(\boldsymbol{\beta})$ converges in probability to a positive definite matrix, say $S(\boldsymbol{\beta})$, we have with $\Sigma = S^{-1}$

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n \; - \; \boldsymbol{\beta}) \xrightarrow{\text{D}} \mathsf{N}(0, \Sigma(\boldsymbol{\beta})).$$

2.2.2 Logistic regression and 2×2 tables

Now, let us consider the following situation. The response Y describes the occurrence of a disease; we are interested in the dependence of this occurrence on the presence of a so-called exposure, which is characterized by a variable X. That is, also the covariate is an indicator. Usually such a problem is described by a 2×2 table:

	$X = 1$	$X = 0$	
$Y = 1$	H_{11}	H_{10}	H_{1+}
$Y = 0$	H_{01}	H_{00}	H_{0+}
	H_{+1}	H_{+0}	n

with

$$H_{jk} = \sum_{i=1} \mathbb{1}(Y_i = j, X_i = k).$$

We introduce the following probabilities

$$p_{11} = \mathsf{P}(Y = 1, X = 1), \quad p_{10} = \mathsf{P}(Y = 1, X = 0), \quad p_{01} = \mathsf{P}(Y = 0, X = 1)$$

and

$$p_{00} = \mathsf{P}(Y = 0, X = 0) \quad \text{with} \quad p_{11} + p_{10} + p_{01} + p_{00} = 1.$$

The odds are

$$\psi_0 = \frac{\mathsf{P}(Y = 1|X = 0)}{\mathsf{P}(Y = 0|X = 0)} = \frac{p_{10}}{p_{00}} \qquad \psi_1 = \frac{\mathsf{P}(Y = 1|X = 1)}{\mathsf{P}(Y = 0|X = 1)} = \frac{p_{11}}{p_{01}}$$

and the odds ratio and its logarithm are given by

$$\lambda = \frac{\psi_1}{\psi_0} = \frac{p_{11}p_{00}}{p_{10}p_{01}}$$

and

$$\beta^* = \log \lambda = \log p_{11} + \log p_{00} - \log p_{10} - \log p_{01}.$$

In other words, now also the distribution of the covariate is modeled. We have a multinomial distribution (with three unknown parameters) and we can derive the maximum likelihood estimators for the p_{jk}'s. The likelihood function is given by

$$L(\boldsymbol{p}) \propto p_{11}^{H_{11}}\, p_{10}^{H_{10}}\, p_{01}^{H_{01}}\, (1 - p_{11} - p_{10} - p_{01})^{H_{00}},$$

the log likelihood is

$$\ell(\boldsymbol{p}) = H_{11}\log p_{11} + H_{10}\log p_{10} + H_{01}\log p_{01} + H_{00}\log(1 - p_{11} - p_{10} - p_{01}).$$

It is easy to show that the maximum likelihood estimators (MLE) for the probabilities are given by

$$\widehat{p}_{jkn} = \frac{H_{jk}}{n}$$

and consequently, λ and β^* are estimated by

$$\widehat{\lambda}_n = \frac{H_{11}H_{00}}{H_{10}H_{01}} \qquad \widehat{\beta}^*_n = \log\frac{H_{11}H_{00}}{H_{10}H_{01}},$$

respectively.
The asymptotic properties of $\widehat{\lambda}_n$ and $\widehat{\beta}^*_n$ can be derived by using the Fisher information in the multinomial model or directly by the delta method. We obtain for the Fisher information included in one observation (Y_i, X_i)

$$i(\boldsymbol{p}) = \begin{pmatrix} p_{11}^{-1} + p_{00}^{-1} & p_{00}^{-1} & p_{00}^{-1} \\ p_{00}^{-1} & p_{10}^{-1} + p_{00}^{-1} & p_{00}^{-1} \\ p_{00}^{-1} & p_{00}^{-1} & p_{01}^{-1} + p_{00}^{-1} \end{pmatrix}. \tag{2.3}$$

It follows that

$$\sqrt{n}(\widehat{\boldsymbol{p}}_n - \boldsymbol{p}) \xrightarrow{\text{D}} \mathsf{N}_3(0, A(\boldsymbol{p}))$$

with

$$A(\boldsymbol{p}) = i^{-1}(\boldsymbol{p}) = \begin{pmatrix} p_{11}(1-p_{11}) & -p_{11}p_{10} & -p_{11}p_{01} \\ -p_{11}p_{10} & p_{10}(1-p_{10}) & -p_{10}p_{01} \\ -p_{11}p_{01} & -p_{10}p_{01} & p_{01}(1-p_{01}) \end{pmatrix}.$$

Let us apply the delta method: Set $\beta^* = g(\boldsymbol{p})$

and let ∇g be the vector of the partial derivatives of g. By the formula $\sigma^2(\boldsymbol{p}) = \nabla g(\boldsymbol{p})^T A(\boldsymbol{p}) \nabla g(\boldsymbol{p})$, one obtains

$$\sigma^2(\boldsymbol{p}) = p_{11}^{-1} + p_{10}^{-1} + p_{01}^{-1} + p_{00}^{-1}.$$

And by the central limit theorem it follows

$$\sqrt{n}(\widehat{\beta}^*_n - \beta^*) \xrightarrow{\text{D}} \mathsf{N}(0, \sigma^2(\boldsymbol{p})).$$

It is obvious that the asymptotic variance of $\widehat{\beta}_n^*$ is estimated by its MLE

$$\widehat{\sigma}_n^2/n = H_{11}^{-1} + H_{10}^{-1} + H_{01}^{-1} + H_{00}^{-1}. \tag{2.4}$$

Now, let us consider the 2×2 table under the viewpoint of the logistic regression: We have by (2.1)

$$\pi(0, \boldsymbol{\beta}) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \qquad \pi(1, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}.$$

The odds ratio has now the form

$$\lambda = \frac{\pi(1, \boldsymbol{\beta})}{1 - \pi(1, \boldsymbol{\beta})} \frac{1 - \pi(0, \boldsymbol{\beta})}{\pi(0, \boldsymbol{\beta})} = \exp(\beta_1),$$

i.e. $\beta^* = \beta_1$.
The score equations for the derivation of the estimators $\widehat{\beta}_{0n}$ and $\widehat{\beta}_{1n}$ have the form

$$\begin{aligned}
(1) \quad & \sum_{\substack{i=1 \\ x_i=1}} (Y_i - \pi(1, \boldsymbol{\beta})) + \sum_{\substack{i=1 \\ x_i=0}} (Y_i - \pi(0, \boldsymbol{\beta})) \\
& = H_{11} - H_{+1}\pi(1, \boldsymbol{\beta}) + H_{10} - H_{+0}\pi(0, \boldsymbol{\beta}) = 0 \qquad (2.5) \\
(2) \quad & \sum_{\substack{i=1 \\ x_i=1}} (Y_i - \pi(1, \boldsymbol{\beta})) = H_{11} - H_{+1}\pi(1, \boldsymbol{\beta}) = 0.
\end{aligned}$$

By straightforward computations, one can show that

$$\widehat{\beta}_{0n} = \log \frac{H_{10}}{H_{00}} \quad \text{and} \quad \widehat{\beta}_{1n} = \log \frac{H_{11} H_{00}}{H_{10} H_{01}}$$

are solutions of (2.5).
Thus, both approaches lead to the same estimator for $\beta^* = \beta_1$. The conditional observed information matrix is given by $\mathcal{I}_n$ with the elements

$$\begin{aligned}
\mathcal{I}_{n\,11}(\boldsymbol{\beta}) &= H_{+0}\pi(0, \boldsymbol{\beta})(1 - \pi(0, \boldsymbol{\beta})) + H_{+1}\pi(1, \boldsymbol{\beta})(1 - \pi(1, \boldsymbol{\beta})) \\
\mathcal{I}_{n\,12}(\boldsymbol{\beta}) &= H_{+1}\pi(1, \boldsymbol{\beta})(1 - \pi(1, \boldsymbol{\beta})) \\
\mathcal{I}_{n\,22}(\boldsymbol{\beta}) &= H_{+1}\pi(1, \boldsymbol{\beta})(1 - \pi(1, \boldsymbol{\beta}))
\end{aligned}$$

and $n^{-1}\mathcal{I}_n(\boldsymbol{\beta})$ converges with probability one to the positive definite matrix $S(\boldsymbol{\beta})$

$$\begin{aligned} S_{11}(\boldsymbol{\beta}) &= p_{+0}\pi(0,\boldsymbol{\beta})(1-\pi(0,\boldsymbol{\beta})) + p_{+1}\pi(1,\boldsymbol{\beta})(1-\pi(1,\boldsymbol{\beta})) \\ S_{12}(\boldsymbol{\beta}) &= p_{+1}\pi(1,\boldsymbol{\beta})(1-\pi(1,\boldsymbol{\beta})) \\ S_{22}(\boldsymbol{\beta}) &= p_{+1}\pi(1,\boldsymbol{\beta})(1-\pi(1,\boldsymbol{\beta})). \end{aligned} \tag{2.6}$$

The inverse of S gives the asymptotic variance, and we obtain for the element of interest

$$\Sigma_{22}(\boldsymbol{\beta}) = \sigma^2(\boldsymbol{p}).$$

2.3 Estimators in case-cohort models with discrete covariates

Now we study the estimation in the case-cohort model. We begin with the special case of a 2×2 table and translate these results into the notation of the logistic regression. This we take as a starting point for the investigation of the inference in logistic regression models with discrete covariates.
Let us describe the selection procedure: The subcohort is selected at random from the entire cohort. Let us introduce the indicator V, i.e., $V_i = 1$ if the individual i is an element of the subcohort, and $V_i = 0$ otherwise. We assume that $\mathsf{P}(V_i = 1) = \alpha$. The random variables V_i, Y_i and X_i are independent.
As described in the introduction, the covariate X is observed for all individuals with $Y_i = 1$ and for all individuals in the subcohort.

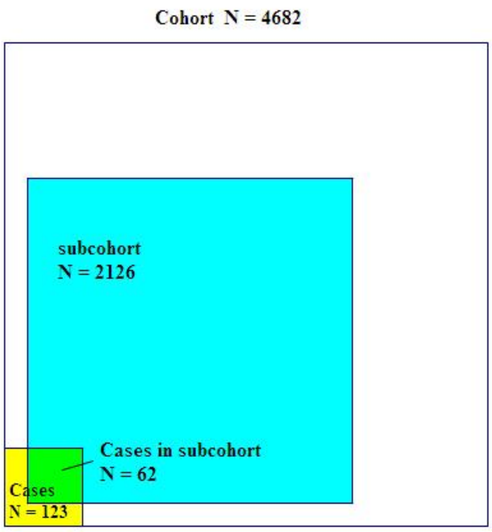
For example, Cologne et al. (2012) illustrated the initial design of the immunogenome and cancer case-cohort study using simple random sampling. Numbers of subjects in the immunogenome study cohort is 4682 and the subcohort was collected using a sampling fraction of 0.5 to reduce genotyping effort by about one-half while retaining most of the full-cohort power.

In Table 2.1, the number of lung cancer cases and subcohort sizes are presented. The proportional of numbers of observations is displayed in Figure 2.1. The case-cohort set is the union of cases set and subcohort set, i.e. the total number of case-cohort observations is 2187.
In this study, they collected covariates as follows: city of residence, gender, year of birth, smoking frequency, whole-body radiation dose, an indicator of EGFR gene CA repeat length < 38 and the product of radiation dose and EGFR CA repeat length indicator (see Cologne et al. (2012)).

Table 2.1: Numbers of individuals in the immunogenome study cohort and case-cohort.

	Subcohort	Non-subcohort	Total
Number of lung cancer cases	62	61	123
Number of non-lung cancer cases	2064	2495	4559
Total	2126	2556	4682

Figure 2.1: Conceptual illustration of the immunogenome and cancer case-cohort design. The proportional of numbers of observations are represented by areas of different colors.



2.3.1 The maximum likelihood estimator in 2×2-tables and their properties

For the binary covariate model the case-cohort design leads to the following frequency table
with

$$D_1 = \sum_{i=1}^{n} Y_i X_i \quad D_0 = \sum_{i=1}^{n} Y_i (1 - X_i)$$

$$R_1 = \sum_{i=1}^{n} (1 - Y_i) X_i V_i \quad R_0 = \sum_{i=1}^{n} (1 - Y_i)(1 - X_i) V_i$$

	$Y=1$		$Y=0$		
	$X=1$	$X=0$	$X=1$	$X=0$	
$V=1$	K_1	K_0	R_1	R_0	$K+R=N_0$
$V=0$	D_1-K_1	D_0-K_0	$n-R-D$		$n-(K+R)$
	D_1	D_0	$n-D$		n

$$K_1 = \sum_{i=1}^{n} Y_i X_i V_i \quad K_0 = \sum_{i=1}^{n} Y_i (1 - X_i) V_i$$

$$D = D_0 + D_1, \quad R = R_0 + R_1, \quad K = K_0 + K_1, \quad N_0 = \sum_{i=1}^{n} V_i.$$

Let us derive the MLE for the probabilities p_{jk} and the parameter β: The contribution to the likelihood of an individual inside the subcohort is given by

$$p_{11}^{Y_i X_i V_i}\, p_{10}^{Y_i(1-X_i)V_i}\, p_{01}^{(1-Y_i)X_i V_i}\, (1 - p_{11} - p_{10} - p_{01})^{(1-Y_i)(1-X_i)V_i}$$

and of an individual not in subcohort

$$p_{11}^{Y_i X_i (1-V_i)}\, p_{10}^{Y_i(1-X_i)(1-V_i)}\, (1 - p_{11} - p_{10})^{(1-Y_i)(1-V_i)}.$$

Thus the likelihood and the log likelihood have the form

$$L(p_{11}, p_{10}, p_{01}) = p_{11}^{D_1}\, p_{10}^{D_0}\, p_{01}^{R_1} (1 - p_{11} - p_{10})^{(n-D-R)} (1 - p_{11} - p_{10} - p_{01})^{R_0}$$

$$\begin{aligned}\ell(p_{11}, p_{10}, p_{01}) \;=\; & D_1 \log p_{11} \;+\; D_0 \log p_{10} \;+\; R_1 \log p_{01} \\ & + R_0 \log(1 - p_{11} - p_{10} - p_{01}) \\ & + (n - D - R) \log(1 - p_{11} - p_{10}).\end{aligned}$$

The score equalities are

$$\begin{aligned}\frac{D_1}{p_{11}} - \frac{R_0}{1 - p_{11} - p_{10} - p_{01}} - \frac{n - D - R}{1 - p_{11} - p_{10}} &= 0 \\ \frac{D_0}{p_{10}} - \frac{R_0}{1 - p_{11} - p_{10} - p_{01}} - \frac{n - D - R}{1 - p_{11} - p_{10}} &= 0 \\ \frac{R_1}{p_{01}} - \frac{R_0}{1 - p_{11} - p_{10} - p_{01}} &= 0.\end{aligned}$$

Notice, for $V_i = 1$ for all i, we have $D_1 = H_{11}$, $D_0 = H_{10}$, $R_1 = H_{01}$, $R_0 = H_{00}$ and $n = D + R$, that is, the equations are the same as in the full cohort model. Solving these equations leads to following MLE

$$\widehat{p}_{11n} = \frac{D_1}{n}, \qquad \widehat{p}_{10n} = \frac{D_0}{n}, \qquad \widehat{p}_{01n} = \frac{(n-D)R_1}{nR}.$$

and

$$\widehat{p}_{00n} = \frac{(n-D)R_0}{nR},$$

and therefore

$$\widehat{\lambda}_n = \frac{D_1\,R_0}{D_0\,R_1} \qquad \text{and } \widehat{\beta}^*_n = \log \widehat{\lambda}_n.$$

Using another approach, this estimator was proposed by Prentice (1986).

2.3.1.1 Properties of the estimators

Now, let us consider the asymptotic properties of the estimators $\widehat{p}_{jkn}$: For this aim we derive the Fisher information matrix for one observation triple (Y_i, X_i, V_i). Set $p = \mathsf{P}(Y_i = 0) = 1 - p_{11} - p_{10}$, and for simplicity of notation we write $p_{00} = 1 - p_{11} - p_{10} - p_{01}$. Computing

$$-\mathsf{E}\,\frac{\partial^2 \ell(\boldsymbol{p})}{\partial \boldsymbol{p} \partial \boldsymbol{p}^T}/n$$

we obtain the Fisher information

$$i(\boldsymbol{p}, \alpha) = \begin{pmatrix} p_{11}^{-1} + \alpha p_{00}^{-1} + (1-\alpha)p^{-1} & \alpha p_{00}^{-1} + (1-\alpha)p^{-1} & \alpha p_{00}^{-1} \\ \alpha p_{00}^{-1} + (1-\alpha)p^{-1} & p_{10}^{-1} + \alpha p_{00}^{-1} + (1-\alpha)p^{-1} & \alpha p_{00}^{-1} \\ \alpha p_{00}^{-1} & \alpha p_{00}^{-1} & \alpha p_{01}^{-1} + \alpha p_{00}^{-1} \end{pmatrix}.$$

For $\alpha = 1$ we obtain the matrix in (2.3).
Based on the information matrix we prove the following theorem on the asymptotic normality of the estimators $\widehat{\boldsymbol{p}}_n$ and $\widehat{\beta}^*_n$:

Theorem 2.1. *Under the case-cohort design introduced above the estimators $\widehat{\boldsymbol{p}}_n$ and $\widehat{\beta}^*_n$ are consistent and asymptotically normal with*

$$\sqrt{n}(\widehat{\boldsymbol{p}}_n - \boldsymbol{p}) \xrightarrow{\text{D}} \mathsf{N}_3(0, \tilde{A}(\boldsymbol{p}, \alpha))$$

with

$$\tilde{A}(\boldsymbol{p},\alpha) = i^{-1}(\boldsymbol{p},\alpha) =$$
$$\begin{pmatrix} p_{11}(1-p_{11}) & -p_{11}p_{10} & -p_{11}p_{01} \\ -p_{11}p_{10} & p_{10}(1-p_{10}) & -p_{10}p_{01} \\ -p_{11}p_{01} & -p_{10}p_{01} & p^{-1}p_{01}(\alpha^{-1}p_{00}+p_{01}(1-p)) \end{pmatrix}$$

and

$$\sqrt{n}(\widehat{\beta}^*_n - \beta^*) \xrightarrow{\text{D}} \mathsf{N}(0,\sigma^2(\boldsymbol{p},\alpha))$$

where

$$\sigma^2(\boldsymbol{p},\alpha) = \frac{1}{p_{11}} + \frac{1}{p_{10}} + \frac{1}{\alpha}\left(\frac{1}{p_{01}} + \frac{1}{p_{00}}\right).$$

Proof. Since $n^{-1}D_j$ and $n^{-1}R_j$, $j = 0, 1$ are sums of i.i.d. random variables the consistency of the estimators follows from the law of large numbers and the continuous mapping theorem.
The limit statement for the vector of the probabilities, i.e., the computation of the variance of the limiting distribution, follows by computing the inverse $i^{-1}(\boldsymbol{p},\alpha)$. Finally, the application of the delta method gives the limit distribution of the log odds ratio. □

An estimator for the variance in the case-cohort design is given by replacing $\sigma^2(\boldsymbol{p},\alpha)$ the unknown probabilities by their MLE. This leads to the following estimator

$$\widehat{\sigma}^2_n(\alpha)/n = \frac{1}{D_1} + \frac{1}{D_0} + \frac{1}{\alpha}\frac{R^2}{(n-D)R_1R_0}. \tag{2.7}$$

Note that in the case $V_i = 1$ for all i and $\alpha = 1$ this estimator coincides with estimator (2.4). Furthermore, this estimator is consistent.
Prentice (1986) proposed also an estimator for the asymptotic variance in the 2×2 table model. This estimator has the following form:

$$\widehat{\sigma}^2_{n\text{prent}}/n = \frac{1}{D_1} + \frac{1}{D_0} + \frac{1}{R_1} + \frac{1}{R_0}.$$

Prentice does not give an expression for the asymptotic variance. It is not difficult to see that the Prentice estimator is also consistent, however it is not the MLE.

In Section (2.3.1.2) we compare both estimators by simulations.

We conclude this section with the investigation of the case-cohort design under the logistic regression aspect. To handle the missing observations of the covariates outside the subcohort for the response $Y_i = 0$ we will make an assumption about the distribution of the X_i. We assume

$$\mathsf{P}(X_i = 1) = f_1 := f \qquad \mathsf{P}(X_i = 0) = f_0 = 1 - f.$$

The likelihood function is

$$L(\boldsymbol{\beta}, f) = \prod_{i=1}^{n} \pi(X_i, \boldsymbol{\beta})^{Y_i} f_{X_i}^{Y_i} (1 - \pi(X_i, \boldsymbol{\beta}))^{(1-Y_i)V_i} f_{X_i}^{(1-Y_i)V_i} p(\boldsymbol{\beta}, f)^{(1-Y_i)(1-V_i)}$$

with

$$p(\boldsymbol{\beta}, f) = \mathsf{P}(Y_i = 0) = f(1 - \pi(1, \boldsymbol{\beta})) + (1 - f)(1 - \pi(0, \boldsymbol{\beta})).$$

and the log likelihood function

$$\begin{aligned}
&\ell(\boldsymbol{\beta}, f) \\
&= \sum_{i=1}^{n} \Big[Y_i X_i \log \pi(1, \boldsymbol{\beta}) + Y_i(1 - X_i) \log \pi(0, \boldsymbol{\beta}) \\
&\quad + Y_i X_i \log f + Y_i(1 - X_i) \log(1 - f) \\
&\quad + V_i(1 - Y_i) X_i \log(1 - \pi(1, \boldsymbol{\beta})) + V_i(1 - Y_i)(1 - X_i) \log(1 - \pi(0, \boldsymbol{\beta})) \\
&\quad + V_i(1 - Y_i) X_i \log f + V_i(1 - Y_i)(1 - X_i) \log(1 - f) \\
&\quad + (1 - Y_i)(1 - V_i) \log \left(f(1 - \pi(1, \boldsymbol{\beta})) + (1 - f)(1 - \pi(0, \boldsymbol{\beta})) \right) \Big] \\
&= D_1 \log \pi(1, \boldsymbol{\beta}) + D_0 \log \pi(0, \boldsymbol{\beta}) \\
&\quad + D_1 \log f + D_0 \log(1 - f) \\
&\quad + R_1 \log(1 - \pi(1, \boldsymbol{\beta})) + R_0 \log(1 - \pi(0, \boldsymbol{\beta})) \\
&\quad + R_1 \log f + R_0 \log(1 - f) \\
&\quad + (n - D - R) \log \left(f(1 - \pi(1, \boldsymbol{\beta})) + (1 - f)(1 - \pi(0, \boldsymbol{\beta})) \right). \qquad (2.8)
\end{aligned}$$

The score equations are given by

$$u_j(\boldsymbol{\beta}, f) = 0 \qquad j = 1, 2, 3$$

with

$$
\begin{aligned}
u_1(\boldsymbol{\beta}, f) =& D_1 - (D_1 + R_1)\pi(1, \boldsymbol{\beta}) + D_0 - (D_0 + R_0)\pi(0, \boldsymbol{\beta}) \\
& - (n - D - R) \frac{f\pi(1, \boldsymbol{\beta})(1 - \pi(1, \boldsymbol{\beta})) + (1 - f)\pi(0, \boldsymbol{\beta})(1 - \pi(0, \boldsymbol{\beta}))}{p(\boldsymbol{\beta}, f)} \\
u_2(\boldsymbol{\beta}, f) =& D_1 - (D_1 + R_1)\pi(1, \boldsymbol{\beta}) - (n - D - R) \frac{f(1 - \pi(1, \boldsymbol{\beta}))\pi(1, \boldsymbol{\beta})}{p(\boldsymbol{\beta}, f)} \\
u_3(\boldsymbol{\beta}, f) =& \frac{D_1 + R_1}{f} - \frac{D_0 + R_0}{1 - f} + (n - D - R) \frac{\pi(0, \boldsymbol{\beta}) - \pi(1, \boldsymbol{\beta})}{p(\boldsymbol{\beta}, f)}.
\end{aligned}
$$

Note, for $V_i = 1$ for all i, $D_1 = H_{11}$ $D_1 + R_1 = H_{+1}$ and $n - D - R = 0$. Thus, the first two equations are equations (2.5).
Solving the likelihood equation for β_0, β^* and f leads via the estimators

$$
\widehat{\pi}(1, \boldsymbol{\beta}) = \frac{D_1 R}{D_1 R + (n - D)R_1} \quad \text{and} \quad \widehat{\pi}(0, \boldsymbol{\beta}) = \frac{D_0 R}{D_0 R + (n - D)R_0}
$$

to

$$
\widehat{\beta}_n^* = \log \frac{D_1 \, R_0}{D_0 \, R_1}, \qquad \widehat{\beta}_{0n} = \frac{D_0 R}{(n - D)R_0}
$$

and

$$
\widehat{f}_n = \frac{D_1 R + (n - D)R_1}{nR}.
$$

By taking into account the distribution of the covariates in the likelihood, we obtain the same results as in the binary model.

2.3.1.2 Simulations

We carry out simulation studies to assess the quality of the estimator depending on the value α and to compare the proposed variance estimator of the estimator $\widehat{\beta_n^*}$ as (2.7) with the theoretic value and the Prentice estimator. We fix the parameter $p_{11} = 0.2, p_{10} = 0.4$ and $p_{01} = 0.2$, that is we have the true value

$$
\beta^* = \log \lambda = \log(\frac{p_{11} p_{00}}{p_{01} p_{10}}) = -0.6931472.
$$

For each configuration, we simulated 3000 full cohort samples with sample size $n = 500$ objects corresponding to the parameters p_{11}, p_{10} and p_{01}, we have a multinomial model with 4 outcomes. It follows that the disease state were generated corresponding to a binary covariate. Moreover, device whether an individual is an element of the subcohort, we generate the indicator V based on $P(V = 1) = \alpha$.

In this section, we first compare the performance of estimators for several different values of α with real value β^*. Table 2.2 presents the simulated estimates results of β^* of subcohort for 4 different values of α ($\alpha = 0.2, 0.5, 0.7$ and 0.9). Furthermore, Figure 2.2 and 2.3 show the estimate of β^* in full cohort and subcohort based on α compare with the real parameter β^*(red line), respectively. As shown in Table 2.2, Figure 2.2 and 2.3, the estimates spread around the true value for several values of α. These results show that the estimates of β^* is reliable.

Table 2.2: Simulation summary statistics for estimation under case-cohort design with different values of α.

α	Mean($\widehat{\beta}_n^*$)	SD($\widehat{\beta}_n^*$)
0.2	-0.6959494	0.3475209
0.5	-0.6974123	0.2362503
0.7	-0.6906599	0.2139101
0.9	-0.6939503	0.1971865

Figure 2.2: Histogram of β^* estimates in full cohort.

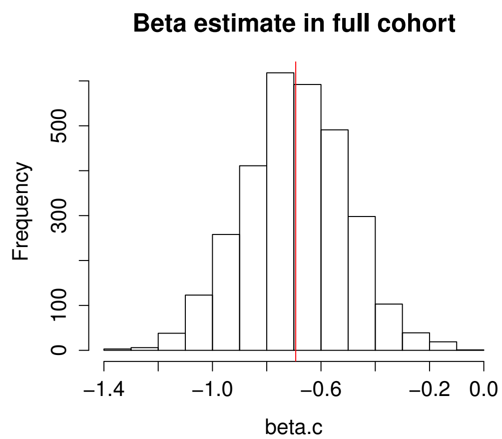
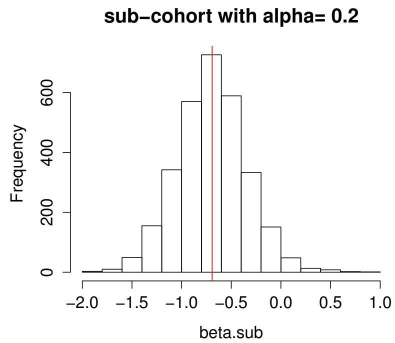
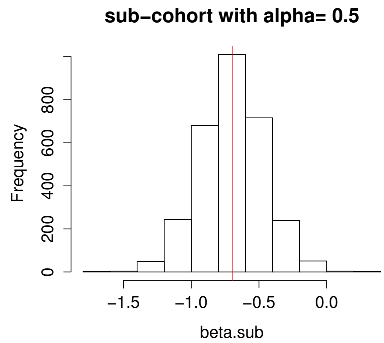


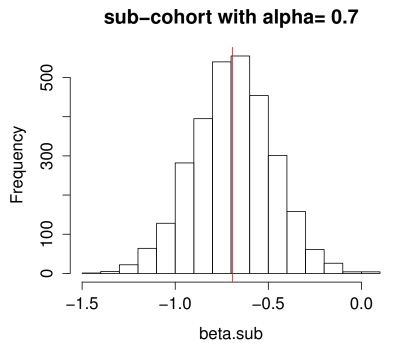
Figure 2.3: Histogram of β^* estimates under the case-cohort design.



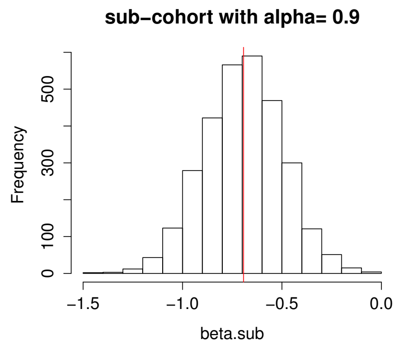
(a) $\alpha = 0.2$



(b) $\alpha = 0.5$



(c) $\alpha = 0.7$



(d) $\alpha = 0.9$

Next, we compare the performance of the proposed estimator and the Prentice estimator. The performance of both variance estimators are assessed via the mean square error (MSE),

$$\text{MSE} = \text{Bias}^2 + \text{Variance}.$$

Under the case-cohort design, we have by Theorem 2.1 asymptotic variance of $\widehat{\beta_n^*}$ as

$$\sigma^2(\boldsymbol{p}, \alpha)/n = (np_{11})^{-1} + (np_{10})^{-1} + \frac{1}{n\alpha}\frac{p_{01} + p_{00}}{p_{01}p_{00}}.$$

The both estimator results are demonstrated in Table 2.3. The empirical bias and the empirical variance of the variance estimates are shown for different values of α. Table 2.3 shows, as expected, that the proposed estimator for variance of the estimator based on the maximum likelihood estimator gives lower empirical bias and empirical variance than the estimator based on the Prentice's estimator.

Table 2.3: The variance estimator of case cohort based on α.

α	$\sigma^2(\boldsymbol{p}, \alpha)/n$	Estimator	Bias	Variance	MSE
0.2	0.115	Prentice	0.0049125590	0.00030782510	0.0003319583
		MLE	0.0032557290	0.00005094680	0.0000061547
0.5	0.055	Prentice	0.0008067732	0.00001278035	0.0000134312
		MLE	0.0006553818	0.00000380526	0.0000042348
0.7	0.044	Prentice	0.0004971469	0.00000356768	0.0000038148
		MLE	0.0004130448	0.00000166494	0.0000018355
0.9	0.037	Prentice	0.0003356457	0.00000122481	0.0000013375
		MLE	0.0003012373	0.00000093929	0.0000010300

2.3.2 The maximum likelihood estimator in logistic regression with discrete covariates

Now we consider the logistic regression model with a one-dimensional discrete covariate taking m values $\xi_1, \ldots, \xi_m$ with

$$f_j = \mathsf{P}(X_i = \xi_j) \qquad j = 1, \ldots, m \qquad \sum_{j=1}^{m} f_j = 1.$$

Since the f_j's are unknown we have a model with $2 + m - 1$ parameters. Generalizing the approach from the previous section, we obtain as likelihood function

$$\begin{aligned} L(\boldsymbol{\beta}, f) &= \prod_{i=1}^{n} L_i(\boldsymbol{\beta}, \boldsymbol{f}; Y_i, X_i, V_i) \\ &= \prod_{i=1}^{n} \pi(X_i, \boldsymbol{\beta})^{Y_i} f_{X_i}^{Y_i} (1 - \pi(X_i, \boldsymbol{\beta}))^{(1-Y_i)V_i} f_{X_i}^{(1-Y_i)V_i} p(\boldsymbol{\beta}, \boldsymbol{f})^{(1-Y_i)(1-V_i)} \end{aligned}$$

with X_i taking values in $\{\xi_1, \ldots, \xi_m\}$ and

$$p(\boldsymbol{\beta}, \boldsymbol{f}) = \sum_{j=1}^{m} (1 - \pi(\xi_j, \boldsymbol{\beta})) f_j.$$

The log likelihood function is

$$\begin{aligned}\ell(\boldsymbol{\beta}, \boldsymbol{f}) &= \sum_{j=1}^{m} \big\{ D_j \log \pi(\xi_j, \boldsymbol{\beta}) + R_j \log(1 - \pi(\xi_j, \boldsymbol{\beta})) \\ &\quad + (D_j + R_j) \log f_j + (n - D - R) \log p(\boldsymbol{\beta}, \boldsymbol{f}) \big\}.\end{aligned}$$

with

$$D_j = \sum_{i=1}^{n} Y_i \mathbb{1}(X_i = \xi_j) \quad \text{and} \quad R_j = \sum_{i=1}^{n} (1 - Y_i) V_i \mathbb{1}(X_i = \xi_j).$$

The elements of the score vectors are given by

$$\begin{aligned}u_1(\boldsymbol{\beta}, \boldsymbol{f}) &= \sum_{j=1}^{m} D_j - \sum_{j=1}^{m} (D_j + R_j) \pi(\xi_j, \boldsymbol{\beta}) + \frac{n - D - R}{p(\boldsymbol{\beta}, \boldsymbol{f})} p_{\beta_0}(\boldsymbol{\beta}, \boldsymbol{f}) \\ u_2(\boldsymbol{\beta}, \boldsymbol{f}) &= \sum_{j=1}^{m} D_j \xi_j - \sum_{j=1}^{m} (D_j + R_j) \pi(\xi_j, \boldsymbol{\beta}) \xi_j + \frac{n - D - R}{p(\boldsymbol{\beta}, \boldsymbol{f})} p_{\beta_1}(\boldsymbol{\beta}, \boldsymbol{f}) \\ u_k(\boldsymbol{\beta}, \boldsymbol{f}) &= \frac{D_{k-2} + R_{k-2}}{f_{k-2}} - \frac{D_m + R_m}{1 - \sum_{s=1}^{m-1} f_s} + \frac{n - D - R}{p(\boldsymbol{\beta}, \boldsymbol{f})} p_{f_{k-2}}(\boldsymbol{\beta}, \boldsymbol{f}), \\ &\qquad k = 3, \ldots, m+1\end{aligned}$$

where

$$p_{\beta_0}(\boldsymbol{\beta}, \boldsymbol{f}) = \frac{\partial p(\boldsymbol{\beta}, \boldsymbol{f})}{\partial \beta_0} = -\sum_{j=1}^{m} \pi(\xi_j, \boldsymbol{\beta})(1 - \pi(\xi_j, \boldsymbol{\beta})) f_j$$

$$p_{\beta_1}(\boldsymbol{\beta}, \boldsymbol{f}) = \frac{\partial p(\boldsymbol{\beta}, \boldsymbol{f})}{\partial \beta_1} = -\sum_{j=1}^{m} \pi(\xi_j, \boldsymbol{\beta})(1 - \pi(\xi_j, \boldsymbol{\beta})) \xi_j f_j$$

$$p_{f_{k-2}}(\boldsymbol{\beta}, \boldsymbol{f}) = \frac{\partial p(\boldsymbol{\beta}, \boldsymbol{f})}{\partial f_{k-2}} = \pi(\xi_m, \boldsymbol{\beta}) - \pi(\xi_{k-2}, \boldsymbol{\beta}).$$

The maximum likelihood estimators $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{f}}_n$ are the solution of

$$u_s(\boldsymbol{\beta}, \boldsymbol{f}) = 0 \qquad s = 1, \ldots, m+1.$$

Example 1. Now, to demonstrate the estimation procedure, let us consider the logistic regression model with a one-dimensional discrete covariate taking

3 different values of covariate ξ_1, ξ_2, ξ_3. The log likelihood function can be derived by

$$\begin{aligned}\ell(\boldsymbol{\beta}, \boldsymbol{f}) &= \sum_{j=1}^{3} \big\{ D_j \log \pi(\xi_j, \boldsymbol{\beta}) + R_j \log(1 - \pi(\xi_j, \boldsymbol{\beta})) \\ &\quad + (D_j + R_j) \log f_j + (n - D - R) \log p(\boldsymbol{\beta}, \boldsymbol{f}) \big\}.\end{aligned}$$

The elements of the score vectors are given by

$$\begin{aligned}u_1(\boldsymbol{\beta}, \boldsymbol{f}) =& D_1(1 - \pi(\xi_1, \boldsymbol{\beta})) + D_2(1 - \pi(\xi_2, \boldsymbol{\beta})) + D_3(1 - \pi(\xi_3, \boldsymbol{\beta})) \\ & - R_1\pi(\xi_1, \boldsymbol{\beta}) - R_2\pi(\xi_2, \boldsymbol{\beta}) - R_3\pi(\xi_3, \boldsymbol{\beta}) + \frac{n - D - R}{p(\boldsymbol{\beta}, \boldsymbol{f})} p_{\beta_0}(\boldsymbol{\beta}, \boldsymbol{f}) \\ u_2(\boldsymbol{\beta}, \boldsymbol{f}) =& D_1\xi_1(1 - \pi(\xi_1, \boldsymbol{\beta})) + D_2\xi_2(1 - \pi(\xi_2, \boldsymbol{\beta})) + D_3\xi_3(1 - \pi(\xi_3, \boldsymbol{\beta})) \\ & - R_1\xi_1\pi(\xi_1, \boldsymbol{\beta}) - R_2\xi_2\pi(\xi_2, \boldsymbol{\beta}) - R_3\xi_3\pi(\xi_3, \boldsymbol{\beta}) + \frac{n - D - R}{p(\boldsymbol{\beta}, \boldsymbol{f})} p_{\beta_1}(\boldsymbol{\beta}, \boldsymbol{f}) \\ u_3(\boldsymbol{\beta}, \boldsymbol{f}) =& \frac{D_1 + R_1}{f_1} - \frac{D_3 + R_3}{f_3} + \frac{n - D - R}{p(\boldsymbol{\beta}, \boldsymbol{f})} (\pi(\xi_3, \boldsymbol{\beta}) - \pi(\xi_1, \boldsymbol{\beta})) \\ u_4(\boldsymbol{\beta}, \boldsymbol{f}) =& \frac{D_2 + R_2}{f_2} - \frac{D_3 + R_3}{f_3} + \frac{n - D - R}{p(\boldsymbol{\beta}, \boldsymbol{f})} (\pi(\xi_3, \boldsymbol{\beta}) - \pi(\xi_2, \boldsymbol{\beta}))\end{aligned}$$

Figure 2.4 shows the function $\pi(x)$ with several x under $\beta_0 = -3$ and $\beta_1 = 1.5$.

For illustration, we consider 3 different values of covariate ($\xi_1 = 2, \xi_2 = 3, \xi_3 = 4$) and fix $\beta_0 = -3$ and $\beta_1 = 1.5$. We generate full cohort samples with sample size $n = 1000$ objects corresponding to the parameters f_1, f_2 and f_3, therefore we have a multinomial model with 3 outcomes. We further generate the indicator V based on $P(V = 1) = \alpha$ for the individuals of subcohort.
We calculate the average and standard deviation of estimates based on 500 repeated samples. Then, the maximum likelihood estimates results are summarized in Table 2.4 under several $f_j = \mathsf{P}(X_i = \xi_j), \quad j = 1, \ldots, 3$. The results show that the estimate values are around the true values for several probabilities f_j and different values of α. □

Figure 2.4: The plot of function $\pi(x)$ under $\beta_0 = -3$ and $\beta_1 = 1.5$ with several x.

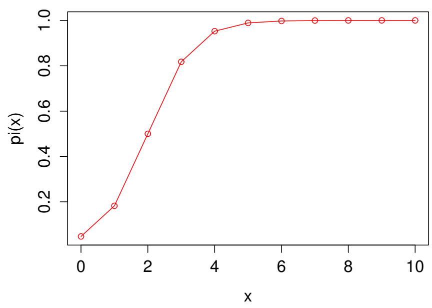


Table 2.4: The estimates results of logistic regression with 3 different values of covariate.

α	f_1, f_2	$Mean(\widehat{\beta_0})(SD)$	$Mean(\widehat{\beta_1})(SD)$	$Mean(\widehat{f_1})(SD)$	$Mean(\widehat{f_2})(SD)$
0.2	0.3333, 0.3333	-3.0617(0.6689)	1.5298(0.2759)	0.3340(0.0216)	0.3330(0.0191)
	0.2, 0.4	-3.0561(0.7058)	1.5278(0.2693)	0.2017(0.0192)	0.3994(0.0181)
	0.4, 0.2	-3.0876(0.6669)	1.5442(0.2911)	0.3993(0.0202)	0.2012(0.0144)
	0.4, 0.4	-3.0848(0.6242)	1.5414(0.2644)	0.4019(0.0220)	0.3983(0.0196)
	0.2, 0.3	-3.0688(0.6949)	1.5331(0.2643)	0.2018(0.0188)	0.2987(0.0163)
	0.3, 0.2	-3.0781(0.6225)	1.5371(0.2557)	0.3011(0.0192)	0.1988(0.0142)
	0.5, 0.3	-3.0386(0.6564)	1.5255(0.2911)	0.5009(0.0218)	0.2992(0.0184)
	0.3, 0.5	-3.0859(0.7289)	1.5403(0.2957)	0.3017(0.0220)	0.4983(0.0211)
	0.2, 0.5	-3.0403(0.7092)	1.5210(0.2690)	0.2015(0.0189)	0.4984(0.0190)
	0.5, 0.2	-3.0968(0.6887)	1.5523(0.3085)	0.5020(0.0203)	0.1984(0.0150)
0.5	0.3333, 0.3333	-3.0185(0.4375)	1.5084(0.1750)	0.3334(0.0177)	0.3339(0.0173)
	0.2, 0.4	-3.0198(0.4616)	1.5102(0.1715)	0.2008(0.0151)	0.4004(0.0168)
	0.4, 0.2	-3.0433(0.4103)	1.5217(0.1699)	0.4009(0.0170)	0.1995(0.0132)
	0.4, 0.4	-3.0130(0.4239)	1.5081(0.1727)	0.4005(0.0174)	0.3996(0.0167)
	0.2, 0.3	-3.0228(0.4490)	1.5113(0.1650)	0.2009(0.0151)	0.2995(0.0155)
	0.3, 0.2	-3.0331(0.4135)	1.5150(0.1617)	0.3006(0.0163)	0.1993(0.0130)
	0.5, 0.3	-2.9919(0.4189)	1.5019(0.1795)	0.5007(0.0178)	0.2996(0.0158)
	0.3, 0.5	-3.0414(0.4628)	1.5186(0.1813)	0.3008(0.0172)	0.4994(0.0176)
	0.2, 0.5	-3.0091(0.4783)	1.5059(0.1788)	0.2006(0.0149)	0.4994(0.0170)
	0.5, 0.2	-3.0076(0.4095)	1.5088(0.1755)	0.5011(0.0173)	0.1993(0.0138)
0.7	0.3333, 0.3333	-3.0380(0.3920)	1.5160(0.1557)	0.3338(0.0170)	0.3338(0.0169)
	0.2, 0.4	-2.9850(0.4024)	1.4961(0.1447)	0.2004(0.0138)	0.4005(0.0163)
	0.4, 0.2	-3.0118(0.3468)	1.5071(0.1383)	0.4003(0.0156)	0.2000(0.0130)
	0.4, 0.4	-2.9977(0.3887)	1.5011(0.1552)	0.4002(0.0159)	0.3998(0.0157)
	0.2, 0.3	-2.9879(0.3815)	1.4970(0.1355)	0.2004(0.0137)	0.2996(0.0148)
	0.3, 0.2	-3.0216(0.3533)	1.5094(0.1337)	0.3004(0.0157)	0.1997(0.0129)
	0.5, 0.3	-3.0169(0.3415)	1.5125(0.1443)	0.5016(0.0170)	0.2992(0.0149)
	0.3, 0.5	-3.0166(0.3937)	1.5078(0.1499)	0.3002(0.0160)	0.4999(0.0164)
	0.2, 0.5	-3.0068(0.4286)	1.5045(0.1576)	0.2004(0.0138)	0.4997(0.0165)
	0.5, 0.2	-3.0049(0.3476)	1.5072(0.1474)	0.5013(0.0171)	0.1992(0.0136)

$$
\begin{aligned}
J_{n\,11} &= \sum_{j=1}^{m}(D_j+R_j)\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta})) \\
&\quad -\frac{n-D-R}{p(\boldsymbol{\beta},\boldsymbol{f})}\left(p_{\beta_0\beta_0}(\boldsymbol{\beta},\boldsymbol{f})-\frac{p_{\beta_0}^2(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
J_{n\,12} &= \sum_{j=1}^{m}(D_j+R_j)\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))\xi_j \\
&\quad -\frac{n-D-R}{p(\boldsymbol{\beta},\boldsymbol{f})}\left(p_{\beta_0\beta_1}(\boldsymbol{\beta},\boldsymbol{f})-\frac{p_{\beta_0}(\boldsymbol{\beta},\boldsymbol{f})p_{\beta_1}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
J_{n\,22} &= \sum_{j=1}^{m}(D_j+R_j)\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))\xi_j^2 \\
&\quad -\frac{n-D-R}{p(\boldsymbol{\beta},\boldsymbol{f})}\left(p_{\beta_1\beta_1}(\boldsymbol{\beta},\boldsymbol{f})-\frac{p_{\beta_1}^2(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
J_{n\,jk} &= -\frac{n-D-R}{p(\boldsymbol{\beta},\boldsymbol{f})}\left(p_{\beta_{j-1}f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f})-\frac{p_{\beta_{j-1}}(\boldsymbol{\beta},\boldsymbol{f})p_{f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
&\quad j=1,2\ k=3,\ldots,m+1 \\
J_{n\,kk} &= \frac{D_{k-2}+R_{k-2}}{f_{k-2}^2}+\frac{D_m+R_m}{f_m^2} \\
&\quad -\frac{n-D-R}{p(\boldsymbol{\beta},\boldsymbol{f})}\left(p_{f_{k-2}f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f})-\frac{p_{f_{k-2}}^2(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
&\quad k=3,\ldots,m+1 \\
J_{n\,kl} &= \frac{D_m+R_m}{f_m^2} \\
&\quad -\frac{n-D-R}{p(\boldsymbol{\beta},\boldsymbol{f})}\left(p_{f_{k-2}f_{l-2}}(\boldsymbol{\beta},\boldsymbol{f})-\frac{p_{f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f})p_{f_{l-2}}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
&\quad k,l=3,\ldots,m+1, l\neq k.
\end{aligned}
$$

The expectation of $J_n(\boldsymbol{\beta},\boldsymbol{f})$ is the Fisher information $\mathcal{I}_n(\boldsymbol{\beta},\boldsymbol{f})$, the Fisher information with respect to one observation is

$$i(\boldsymbol{\beta},\boldsymbol{f})=\mathcal{I}_n(\boldsymbol{\beta},\boldsymbol{f})/n$$

and has the elements

$$
\begin{aligned}
i_{11}(\boldsymbol{\beta},\boldsymbol{f}) &= \sum_{j=1}^{m} \pi^2(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))f_j + \alpha\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))^2 f_j \\
&\quad -(1-\alpha)\left(p_{\beta_0\beta_0}(\boldsymbol{\beta},\boldsymbol{f}) - \frac{p^2_{\beta_0}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
i_{12}(\boldsymbol{\beta},\boldsymbol{f}) &= \sum_{j=1}^{m} \pi^2(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))\xi_j f_j + \alpha\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))^2 \xi_j f_j \\
&\quad -(1-\alpha)\left(p_{\beta_0\beta_1}(\boldsymbol{\beta},\boldsymbol{f}) - \frac{p_{\beta_0}(\boldsymbol{\beta},\boldsymbol{f})p_{\beta_1}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
i_{22}(\boldsymbol{\beta},\boldsymbol{f}) &= \sum_{j=1}^{m} \pi^2(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))\xi_j^2 f_j + \alpha\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))^2 \xi_j^2 f_j \\
&\quad -(1-\alpha)\left(p_{\beta_1\beta_1}(\boldsymbol{\beta},\boldsymbol{f}) - \frac{p^2_{\beta_1}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
i_{jk}(\boldsymbol{\beta},\boldsymbol{f}) &= -(1-\alpha)\left(p_{\beta_{j-1}f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f}) - \frac{p_{\beta_{j-1}}(\boldsymbol{\beta},\boldsymbol{f})p_{f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
&\quad j=1,2 \; k=3,\ldots,m+1 \\
i_{kk}(\boldsymbol{\beta},\boldsymbol{f}) &= \frac{\pi(\xi_{k-2},\boldsymbol{\beta})+\alpha(1-\pi(\xi_{k-2},\boldsymbol{\beta}))}{f_{k-2}} + \frac{\pi(\xi_m,\boldsymbol{\beta})+\alpha(1-\pi(\xi_m,\boldsymbol{\beta}))}{f_m} \\
&\quad -(1-\alpha)\left(p_{f_{k-2}f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f}) - \frac{p^2_{f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
&\quad k=3,\ldots,m+1 \\
i_{kl}(\boldsymbol{\beta},\boldsymbol{f}) &= +\frac{\pi(\xi_m,\boldsymbol{\beta})+\alpha(1-\pi(\xi_m,\boldsymbol{\beta}))}{f_m} \\
&\quad -(1-\alpha)\left(p_{f_{k-2}f_{l-2}}(\boldsymbol{\beta},\boldsymbol{f}) - \frac{p_{f_{k-2}}(\boldsymbol{\beta},\boldsymbol{f})p_{f_{l-2}}(\boldsymbol{\beta},\boldsymbol{f})}{p(\boldsymbol{\beta},\boldsymbol{f})}\right) \\
&\quad k,l=3,\ldots,m+1, l\neq k.
\end{aligned}
$$

We assume:

A1) The matrices $J_n(\boldsymbol{\beta},\boldsymbol{f})$ and $i(\boldsymbol{\beta},\boldsymbol{f})$ are positive definite.

A2) The matrix $n^{-1}J_n(\boldsymbol{\beta},\boldsymbol{f})$ converges (in probability) to the Fisher information matrix $i(\boldsymbol{\beta},\boldsymbol{f})$.

Theorem 2.2. *For the MLE in the logistic regression under case-cohort design we have*

$$\sqrt{n}\left[\begin{pmatrix} \widehat{\boldsymbol{\beta}}_n \\ \widehat{\boldsymbol{f}}_n \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{f} \end{pmatrix}\right] \xrightarrow{\text{D}} \mathsf{N}_{m+1}(0, C(\boldsymbol{\beta}, \boldsymbol{f}))$$

where $C(\boldsymbol{\beta}, \boldsymbol{f}) = [i(\boldsymbol{\beta}, \boldsymbol{f})]^{-1}$.
Moreover,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\text{D}} \mathsf{N}_2(0, \Sigma(\boldsymbol{\beta}, \boldsymbol{f}))$$

with $\Sigma_{jk} = C_{jk}$ *for* $j, k = 1, 2$.

Proof. Using $\mathsf{E}(n - D - R) = n\mathsf{P}(Y = 0)(1 - \alpha)$ and

$$\mathsf{E}D_j = n\pi(\xi_j, \boldsymbol{\beta})f_j \quad \text{and} \quad \mathsf{E}R_j = n\alpha(1 - \pi(\xi_j, \boldsymbol{\beta}))f_j$$

straightforward computations give

$$\mathsf{E}u_k(\boldsymbol{\beta}, \boldsymbol{f}) = 0.$$

The central limit theorem can be applied to sums of the form

$$\sum_{j=1}^{m} D_j = \sum_{i=1}^{n} Y_i \sum_{j=1}^{m} \mathbb{1}(X_i = \xi_j), \quad \sum_{j=1}^{m} D_j\xi_j = \sum_{i=1}^{n} Y_i \sum_{j=1}^{m} \mathbb{1}(X_i = \xi_j)\xi_j \quad \text{and}$$

$$\sum_{j=1}^{m} R_j = \sum_{i=1}^{n}(1 - Y_i)V_i \sum_{j=1}^{m} \mathbb{1}(X_i = \xi_j).$$

It follows that the score vector is asymptotically normal. The elements of the variance matrix are given by

$$\frac{1}{n}\mathsf{Cov}\left(u_k(\boldsymbol{\beta}, \boldsymbol{f}), u_r(\boldsymbol{\beta}, \boldsymbol{f})\right) \qquad k, r = 1, \ldots, m + 1.$$

Since the underlying distribution fulfills the usual regularity conditions, these elements are the elements of the Fisher matrix $i(\boldsymbol{\beta}, \boldsymbol{f})$. Thus, we have

$$n^{-1/2}\boldsymbol{u}_n(\boldsymbol{\beta}, \boldsymbol{f}) \xrightarrow{\text{D}} \mathsf{N}_{m+1}(0, i(\boldsymbol{\beta}, \boldsymbol{f})).$$

Here $\boldsymbol{u}_n$ denotes the score vector. As in the standard maximum likelihood theory, the estimators behave asymptotically as $\mathcal{I}_n(\boldsymbol{\beta},\boldsymbol{f})^{-1}u_n(\boldsymbol{\beta},\boldsymbol{f})$.

$$\sqrt{n}\left[\begin{pmatrix}\widehat{\boldsymbol{\beta}}_n\\ \widehat{\boldsymbol{f}}_n\end{pmatrix}-\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{f}\end{pmatrix}\right]\approx\sqrt{n}\mathcal{I}_n(\boldsymbol{\beta},\boldsymbol{f})^{-1}u_n(\boldsymbol{\beta},\boldsymbol{f}) \tag{2.9}$$
$$= n^{-1/2}(i(\boldsymbol{\beta},\boldsymbol{f})^{-1}u_n(\boldsymbol{\beta},\boldsymbol{f}))\xrightarrow{\text{D}}\mathsf{N}_{m+1}(0,i(\boldsymbol{\beta},\boldsymbol{f})^{-1}).$$

To justify (2.9) one has to show that the estimators are consistent. Consistency is a consequence of the following: The assumptions A1 and A2 imply that the log likelihood function has a unique maximum. The maximizer converges almost surely to the unique maximizer of the limit $\mathsf{E}L_i(\beta,\boldsymbol{f};V_i,X_i,V_i)$, where the expectation is taken under the true parameter. □

Remark Let us consider the Fisher information matrix $i(\boldsymbol{\beta},\boldsymbol{f})$ for the full cohort, that is for the case $\alpha=1$:

$$i_{11}(\boldsymbol{\beta},\boldsymbol{f})=\sum_{j=1}^{m}\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))f_j,\quad i_{12}(\boldsymbol{\beta},\boldsymbol{f})=\sum_{j=1}^{m}\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))\xi_j f_j,$$
$$i_{22}(\boldsymbol{\beta},\boldsymbol{f})=\sum_{j=1}^{m}\pi(\xi_j,\boldsymbol{\beta})(1-\pi(\xi_j,\boldsymbol{\beta}))\xi_j^2 f_j,\quad i_{jk}(\boldsymbol{\beta},\boldsymbol{f})=0\ \ j=1,2\ \ k=3,\dots,m+1,$$
$$i_{kk}(\boldsymbol{\beta},\boldsymbol{f})=\frac{1}{f_{k-2}}+\frac{1}{f_m}\quad k=3,\dots,m+1,\quad i_{kl}=\frac{1}{f_m}\quad k,l=3,\dots,m+1,l\neq k.$$

We see that this matrix consists of two parts, the upper left 2×2 matrix is the generalization of the matrix (2.6). The lower $(m-1)\times(m-1)$ matrix is an information matrix in a multinomial model as in (2.3).

Corollary 2.3. *Because of continuity $J_n(\boldsymbol{\beta},\boldsymbol{f})/n$ is a consistent estimator of $i(\boldsymbol{\beta},\boldsymbol{f})$, and its inverse for $\Sigma(\boldsymbol{\beta},\boldsymbol{f})$. Replacing now the unknown parameters by the MLE we obtain the following consistent estimator for the standard errors of the estimator $\widehat{\boldsymbol{\beta}}_n$:*

$$\mathrm{se}^2(\widehat{\beta}_{0n})=J_n^{11}(\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{f}}_n)\qquad \mathrm{se}^2(\widehat{\beta}_{1n})=J_n^{22}(\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{f}}_n).$$

Here J_n^{jk} denotes the Element (j,k) of the inverse of the matrix J_n.

Example 2. In this example, we further investigate the comparison between theoretical values and simulation results for the asymptotic variance of parameter estimates in the model with a one-dimensional discrete covariate taking the same values as in Example 1.
By the underlying values, we first compute the Fisher information matrix $i(\boldsymbol{\beta}, \boldsymbol{f})$ and the values of $\Sigma(\boldsymbol{\beta}, \boldsymbol{f})$ can be calculated simply by taking the inverse of this matrix. Thus, we divide the matrix $\Sigma(\boldsymbol{\beta}, \boldsymbol{f})$ by n and take square root of them.
Table 2.5 illustrates the theoretical values and the simulation values for 3 different values of α with several probabilities f_1 and f_2. The simulation values of the variance obtained by Example 1 are in good agreement with the theoretical values. There are a slightly difference between the theoretical values and simulation values in every cases.
Moreover, we can see that the variance will be smaller, for the larger α. In other words, as the number of individual in subcohort grows larger, the variance becomes smaller. □

Table 2.5: The theoretical values and the simulation values of the asymptotic variance with different values of α, f_1 and f_2.

	$\alpha = 0.2$ $f_1 = 0.4,\ f_2 = 0.4$		$\alpha = 0.5$ $f_1 = 0.3333,\ f_2 = 0.3333$		$\alpha = 0.7$ $f_1 = 0.4,\ f_2 = 0.2$	
	Theoretical	Simulation	Theoretical	Simulation	Theoretical	Simulation
$\widehat{\beta}_0$	0.6017	0.6242	0.4092	0.4375	0.3379	0.3468
$\widehat{\beta}_1$	0.2525	0.2644	0.1620	0.1750	0.1361	0.1383
$\widehat{f}_1$	0.0202	0.0220	0.0162	0.0177	0.0161	0.0156
$\widehat{f}_2$	0.0178	0.0196	0.0152	0.0173	0.0128	0.0130

Chapter 3

Cox Models and their extensions: Survey of approaches and results

In survival analysis, we consider the time from an initiating event to an event of interest. We denote this time by $\tilde{T}$. Usually, it is called "survival time" or "lifetime". The big difference compared to other statistical investigations is that one has to wait for the occurrence of events – so when the study ends the event of interest has occurred for some individuals but possible not for all. This situation is described by censoring.
As written in the introduction, we will investigate the effect of covariates $\boldsymbol{X}$ on the survival time. One of the most popular model is the Cox model. The first section of this chapter will summarize the notations and concepts of this model.

3.1 Introduction

We denote the conditional survival function of the continuous random variable $\tilde{T}$ given the covariate $\boldsymbol{X}$ takes the values $\boldsymbol{x}$ by

$$S(t|\boldsymbol{x}) = \mathsf{P}(\tilde{T} > t|\boldsymbol{X} = \boldsymbol{x}).$$

The hazard rate is defined by

$$\lambda(t|\boldsymbol{x}) = \lim_{\triangle t \downarrow 0} \frac{1}{\triangle t} \mathsf{P}(t \leq \tilde{T} < t + \triangle t | \tilde{T} \geq t, \boldsymbol{X} = \boldsymbol{x}). \tag{3.1}$$

As mentioned above, right-censoring is inevitable in most survival studies. In general, we observe instead of copies of the time $\tilde{T}$ i.i.d. $T_i = \min(\tilde{T}_i, C_i)$ and

$$\Delta_i = \mathbb{1}(\tilde{T}_i \leq C_i) = \begin{cases} 1, & \tilde{T}_i \leq C_i \\ 0, & \tilde{T}_i > C_i \end{cases}$$

where C_i are i.i.d. so-called censoring variables. We assume non-informative censoring, i.e. the probability of individuals who drop out of the study should be unrelated to the probability of having the event (conditional on values of the covariates $\boldsymbol{X}$). In other words, time to event and time to censoring are statistically independent on the level of covariates.
We further observe the covariates $\boldsymbol{X}$ as random variables. Methods for estimation and testing in models with fixed covariates are the same, however for the study of their properties one has to take into account the difference between fixed and random $\boldsymbol{X}$.

In the Cox model, the data consists of independent and identically distributed copies $(T_i, \Delta_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$ of $(T, \Delta, \boldsymbol{X})$.

3.2 The Cox model (Proportional hazards model)

The Cox proportional hazards model or Cox model was introduced by Cox (1972), it is the most popular model for analyzing survival data. This model relates an individual subject's hazard function where the effect of covariates are modeled on multiplicative scale and the ratio of the hazards for different individuals is constant over the time.
The Cox model assumes the following form (Cox 1972):

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x}) \tag{3.2}$$

where $\exp(\boldsymbol{\beta}^T \boldsymbol{x}) = \exp(\beta_1 x_1 + \ldots + \beta_p x_p)$ is a hazard ratio, λ_0 is an unknown baseline hazard function (the hazard function for an individual with $\boldsymbol{x} = \boldsymbol{0}$) and $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown parameter to be estimated.
The Cox model (3.2) is often called a *proportional hazards model* (PH model) because the ratio of hazard rate for any two sets of covariates $\boldsymbol{x}$ and $\boldsymbol{x}^*$ is

$$\frac{\lambda(t|\boldsymbol{x})}{\lambda(t|\boldsymbol{x}^*)} = \frac{\lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x})}{\lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x}^*)} = \exp(\boldsymbol{\beta}^T (\boldsymbol{x} - \boldsymbol{x}^*)), \text{ for all } t \geq 0$$

which is a constant over time (so the name of proportional hazards model) and we call this ratio the *risk ratio* or *relative risk*.
In particular, if X_1 is a treatment indicator ($X_1 = 1$ if treatment and $X_1 = 0$ if placebo) and all other covariates have to be fixed the same value, then the *relative risk* (hazard ratio) is the ratio $\lambda(t|\boldsymbol{x})/\lambda(t|\boldsymbol{x}^*) = \exp(\beta_1)$ which is the risk of getting the event for the individual have received the treatment relative to the risk of getting the event for the individual who have received the placebo.

With the hazard rate (3.1), we obtain the survival function S

$$\begin{aligned} S(t|\boldsymbol{x}) &= \exp\big(-\int_0^t \lambda(u|\boldsymbol{x})du\big) \\ &= \exp\big(-\int_0^t \lambda_0(u)\exp(\boldsymbol{\beta}^T\boldsymbol{x})du\big) \\ &= \exp\big(-\Lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{x})\big) \qquad (3.3) \\ &= S_0(t)^{\exp(\boldsymbol{\beta}^T\boldsymbol{x})} \end{aligned}$$

where S_0 is the survival function corresponding to the baseline distribution and $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is a cumulative hazard function.
If one chooses a parametric baseline distribution the model (3.2) is a parametric model; the parameter can be estimated in the usual way by the maximum likelihood (ML) method. If we do not specify the baseline distribution, (3.2) describes a semiparametric model. For this situation, Cox proposed the so-called *partial likelihood*. The term partial likelihood is used because the likelihood formula considers probabilities only for those subjects who fail and does not explicitly consider probabilities for those subjects who are censored. Thus, the likelihood for Cox model does not consider probability for all subjects, and so it is called a "partial likelihood". This will be introduced in Section 3.2.1, where we also outline the estimator and its properties.

3.2.1 Partial likelihood estimator and its properties

Let us present the partial likelihood approach. The observed ordered lifetimes are denoted by $T_{(j)}, j = 1, \ldots, d$ where d is the number of observed (uncensored) lifetimes. We start with the presentation of the partial likelihood method as it was introduced by D. R. Cox and we assume that all lifetimes are distinct, in other words there are no ties, and we have $T_{(1)} < T_{(2)} < \ldots < T_{(d)}$.

Remark 3.1. Many authors provided an alternate partial likelihoods for ties between event times; see Breslow (1974), Efron (1977) and Cox (1972).

Define the **risk set** $\mathcal{R}(t)$ at time t as the set of subjects alive and under observation at time t^-, immediately prior to t:

$$\mathcal{R}(t) = \{i : T_i \geq t\}.$$

For the definition of the estimator we need only the risk set at the lifetimes $T_{(j)}$, however it is defined for all t.
The partial likelihood, based on the hazard function (3.2) as defined by Cox, is given by

$$L_n(\boldsymbol{\beta}) = \prod_{j=1}^{d} \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{X}_{(j)})}{\sum_{i \in \mathcal{R}(T_{(j)})} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)} \tag{3.4}$$

where $\boldsymbol{X}_{(j)}$ denotes the covariate associated with the individual whose lifetime is $T_{(j)}$.
Cox suggested treating the partial likelihood as a regular likelihood function and making inference on $\boldsymbol{\beta}_0$ accordingly. The notation $\boldsymbol{\beta}_0$ denotes the true value of $\boldsymbol{\beta}$. We subsequently obtain the estimate of $\boldsymbol{\beta}_0$, often called maximum partial likelihood estimate (MPLE) by maximizing the partial likelihood.
Let $\ell_n(\boldsymbol{\beta}) = \log L_n(\boldsymbol{\beta})$. We obtain

$$\ell_n(\boldsymbol{\beta}) = \sum_{j=1}^{d} \left[\boldsymbol{\beta}^T \boldsymbol{X}_{(j)} - \log \left\{ \sum_{i \in \mathcal{R}(T_{(j)})} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i) \right\} \right].$$

The MPLE $\widehat{\boldsymbol{\beta}}_n$ is the solution of the system of score equations

$$U_{nk}(\boldsymbol{\beta}) = 0, \quad k = 1, ..., p \tag{3.5}$$

where

$$U_{nk}(\boldsymbol{\beta}) = \frac{\partial \ell_n(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{j=1}^{d} \left[\boldsymbol{X}_{(j)k} - \frac{\sum_{i \in \mathcal{R}(T_{(j)})} \boldsymbol{X}_{ik} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)}{\sum_{i \in \mathcal{R}(T_{(j)})} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)} \right],$$

i.e.

$$\mathbf{U}_n(\boldsymbol{\beta}) = \sum_{j=1}^{d} \left[\boldsymbol{X}_{(j)} - \frac{\sum\limits_{i \in \mathcal{R}(T_{(j)})} \boldsymbol{X}_i \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)}{\sum\limits_{i \in \mathcal{R}(T_{(j)})} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)} \right]. \tag{3.6}$$

Although the estimator $\widehat{\boldsymbol{\beta}}_n$ is not a maximum likelihood estimator methods from the MLE-theory, can be used to show that $\widehat{\boldsymbol{\beta}}_n$ is a consistent and asymptotically normal estimator. Since these properties are a special case of a more general statement proved by Andersen et al. (1993) the exact formulation of the assumptions will be postponed to the next section where the extension of the classical Cox model is considered.
To present the limit theorem, let us derive the observed Fisher information as the negative of the second derivative of the log partial likelihood function. It is denoted by $\mathbf{I}_n(\boldsymbol{\beta})$ and has the elements $I_{ngk}(\boldsymbol{\beta}) = -\dfrac{\partial^2 \ell_n(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_g}$:

$$\mathbf{I}_n(\boldsymbol{\beta}) = \sum_{j=1}^{d} \left[\frac{\sum\limits_{i \in \mathcal{R}(T_{(j)})} \boldsymbol{X}_i^{\otimes 2} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)}{\sum\limits_{i \in \mathcal{R}(T_{(j)})} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)} - \left\{ \frac{\sum\limits_{i \in \mathcal{R}(T_{(j)})} \boldsymbol{X}_i \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)}{\sum\limits_{i \in \mathcal{R}(T_{(j)})} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)} \right\}^{\otimes 2} \right]$$

with $\boldsymbol{X}^{\otimes 2} := \boldsymbol{X}\boldsymbol{X}^T$. The regularly conditions include the assumption that $n^{-1}\mathbf{I}_n(\boldsymbol{\beta})$ is non-singular and converges (in probability) to a positive definite matrix $\boldsymbol{\Sigma}$. The limit statements have then the following form: The estimator is consistent

$$\widehat{\boldsymbol{\beta}}_n \xrightarrow{\text{P}} \boldsymbol{\beta}_0, \quad \text{as} \quad n \to \infty$$

and score vector is asymptotically normal

$$n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0) \xrightarrow{\text{D}} \mathsf{N}(0, \boldsymbol{\Sigma}).$$

From Taylor expansion of the score function around $\boldsymbol{\beta}_0$, we get

$$0 = \mathbf{U}_n(\widehat{\boldsymbol{\beta}}_n) = \mathbf{U}_n(\boldsymbol{\beta}_0) - \mathbf{I}_n(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$$

where $\boldsymbol{\beta}^*$ is on the line segment between $\widehat{\boldsymbol{\beta}}_n$ and $\boldsymbol{\beta}_0$. Then, we obtain

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = (n^{-1}\mathbf{I}_n(\boldsymbol{\beta}^*))^{-1} n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0) \approx \boldsymbol{\Sigma}^{-1} n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0) \tag{3.7}$$

and it follows that

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\text{D}} \mathsf{N}(0, \boldsymbol{\Sigma}^{-1}) \tag{3.8}$$

where $\boldsymbol{\Sigma} = \text{plim}_{n\to\infty} n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0)$.
Since $\widehat{\boldsymbol{\beta}}_n$ is the consistent estimator and $\mathbf{I}_n$ is continuous, $n^{-1}\mathbf{I}_n(\widehat{\boldsymbol{\beta}}_n)$ is a consistent estimator of $\boldsymbol{\Sigma}$.

There are three main procedures for testing hypotheses about the regression parameter $\boldsymbol{\beta}_0$. The first test is the usual test based on the asymptotic normality of MPLE, referred to as *Wald's test*.
A test statistic of the hypothesis $\mathcal{H} : \boldsymbol{\beta}_0 = \boldsymbol{\beta}_{\mathcal{H}}$ is

$$\chi^2_W = (\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{\mathcal{H}})^T \mathbf{I}_n(\widehat{\boldsymbol{\beta}}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{\mathcal{H}})$$

which has a Chi-squared distribution with p degrees of freedom for large samples sizes if $\mathcal{H}$ is true .
The *likelihood ratio test* uses

$$\chi^2_{LR} = 2[\ell(\widehat{\boldsymbol{\beta}}_n) - \ell(\boldsymbol{\beta}_{\mathcal{H}})]$$

which has a large-sample Chi-squared distribution with p degrees of freedom under $\mathcal{H}$.
The *score test* is based on the scores, $\mathbf{U}_n(\boldsymbol{\beta}) = (U_1(\boldsymbol{\beta}), ..., U_p(\boldsymbol{\beta}))^T$,

$$\chi^2_{SC} = \mathbf{U}_n^T(\boldsymbol{\beta}_{\mathcal{H}})\mathbf{I}_n^{-1}(\boldsymbol{\beta}_{\mathcal{H}})\mathbf{U}_n(\boldsymbol{\beta}_{\mathcal{H}})$$

which is asymptotically Chi-squared distributed with p degrees of freedom for large n under the null hypothesis.

3.3 The Cox model with time-dependent covariates

In the Cox model (3.2), the covariates are recorded at the beginning of a study, i.e. values will be fixed throughout the course of the study. In many survival studies, perhaps there are other important covariates whose values change during the period of the study. The covariates which change over time are called *time-dependent covariates*, e.g, smoking status, blood pressure, cholesterol, size of the tumor.

We denote a vector of such covariates, which for the ith individual in the sample by $\boldsymbol{X}_i(t) = (X_{i1}(t), \ldots, X_{ip}(t))^T$, corresponding to the value of these covariates at time t. For time-dependent covariates, we assume that their value is predictable in the sense that the value of the covariate is known at an instant just prior to time t.

The Cox model can be extended to include *time-dependent covariates*

$$\lambda(t; \boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x}(t)). \tag{3.9}$$

Two types of time-dependent covariates are distinguished by Kalbfleisch and Prentice (2002).

The first are "external" or "ancillary" time-dependent covariates whose change in a known way. Non-time dependent covariates considered in the previous section are also external; $X(t)$ is generated externally if it is given at the begin of the study, for example if it is of the form $X(t) = X \cdot g(t)$, where g is a given function. Another type of external covariates are those which are not influenced by the occurrence of the event under study, examples are levels of air pollution, daily temperature as a predictor of survival from a heart attack, etc.

The second type of time-dependent covariates are internal covariates. An "internal" time-dependent covariate is that the change of the covariate over time relates to the characteristics or the behavior of the individual. For example, blood pressure, white blood cell count, cigarette smoking status, disease complications, etc.

The most simple time-dependent covariate is a binary variable that is allowed to change once during follow-up. For example, Andersen and Gill (1982) studied the Cox model with the time-dependent covariate:

$$X(t) = \begin{cases} 1 & \text{if women } i \text{ has been resident in a psychiatric hospital} \\ & \text{during the month } [t - 30 \text{ days}, t), \\ 0 & \text{otherwise.} \end{cases}$$

As another example, Kleinbaum and Klein (2005) considered heart transplant status X_{HT} at time t for an individual identified to have a critical heart condition.

$$X_{HT}(t) = \begin{cases} 1 \text{ if received transplant at some time } t_0 \leq t, \\ 0 \text{ if did not receive transplant by time } t. \end{cases}$$

For an individual receiving a transplant at some time t_0, prior to time t, the covariate X_{HT} is 0 up to t_0, and then remains at 1 thereafter. If the individual has not yet received a transplant by time t, the value of X_{HT} is 0 at time t. That is, the individual who never receives transplant has the covariate $X_{HT} = 0$ for all times during the period of the study.

In the model with time-dependent covariates we observe

$$(T_i, \delta_i, \{\boldsymbol{X}_i(t), t \in [0, T_i]\}) \quad i = 1, \ldots, n.$$

$\{\boldsymbol{X}_i(t), t \in [0, T_i]\}$ is the covariate path of individual i while it is in the study.

If we consider the ratio

$$\frac{\lambda(t; \boldsymbol{x}_1)}{\lambda(t; \boldsymbol{x}_2)} = \frac{\lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x}_1(t))}{\lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x}_2(t))} = \exp(\boldsymbol{\beta}^T (\boldsymbol{x}_1(t) - \boldsymbol{x}_2(t))), \text{ for all } t \geq 0$$

which is certainly not constant, so that the proportional hazard assumption is not satisfied for this model.
The survival function is given by

$$S(t; \boldsymbol{x}) = \exp\big(- \int_0^t \lambda_0(u) \exp(\boldsymbol{\beta}^T \boldsymbol{x}(u)) du \big).$$

The partial likelihood function (3.4) has in the extended Cox model the form

$$L_n(\boldsymbol{\beta}) = \prod_{j=1}^{d} \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{X}_{(j)}(T_{(j)}))}{\sum\limits_{i \in \mathcal{R}(T_{(j)})} \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i(T_{(j)}))} \tag{3.10}$$

As before the MPLE $\widehat{\boldsymbol{\beta}}_n$ can be derived by maximizing the log partial likelihood. However, the computations are more complicated than in the classical model because at each death time we need to know the exact value of the covariate at that death time for all individuals at risk. The management collection and storage of such data are quite difficult to create.
Moreover, the problem of missing observations arise. That means the knowledge of

$$\boldsymbol{X}_i(T_{(1)}), \ldots, \boldsymbol{X}_i(T_{(j)}) \quad \text{for all} \quad i \in \mathcal{R}(T_{(j)}),$$

i.e. of all $\boldsymbol{X}_i(T_{(j)})$ with $T_i \geq T_{(j)}$, is required.

Note that the interpretation of the results has to be done carefully. We have no longer a conditional survival function.
The conditional probability

$$\mathsf{P}(T \geq t|\boldsymbol{X}(t)) = 1,$$

if $\boldsymbol{X}(t)$ is known, the individual must be alive and at risk of failure.

For the application, the results and discussions of extended model and computational issues are given by example in T. M. Therneau and P. M. Grambsch (2000).

The key reference to investigate the asymptotic properties of the maximum partial likelihood estimator in this extended Cox model is the paper of Andersen and Gill (1982) who used the counting process framework and the martingale approach. This approach will be introduced in the following section.
Notice that the classical Cox model with covariates which do not depend on time can be considered as a special case. The properties given in the Section 3.2.1 are a consequence of the results of Andersen and Gill (1982).

3.3.1 Counting process approach to the Cox model

The counting process approach is of great importance in developing the theory of the extended Cox model. Andersen and Gill (1982) extended the Cox model to the counting process framework and the asymptotic properties of the associated estimators for the Cox model have been justified elegantly via martingale theory.
Let us introduce the following counting process notations. The n-dimensional counting process is defined by

$$\mathbf{N}(t) = (N_1(t), \ldots, N_n(t))^T$$

with

$$N_i(t) = \mathbb{1}(T_i \leq t, \Delta_i = 1)$$

which is a counting process of the observed failures for the ith individual, that is $N_i(t)$ jumps only if the lifetime is observed.
Furthermore, we introduce the so-called risk indicator

$$Y_i(t) = \mathbb{1}(T_i \geq t),$$

$Y_i(t)$ indicates whether the subject is at risk (for observing the event of interest). The processes N_i and Y_i are observed in some time interval $[0, \tau]$, $\tau < \infty$.

The *intensity process* of $\mathbf{N}(t)$ is given by

$$\boldsymbol{\alpha}(t) = (\alpha_1(t), \ldots, \alpha_n(t))^T$$

with

$$\alpha_i(t) = Y_i(t)\lambda_i(t) = Y_i(t)\lambda_0(t)\exp(\boldsymbol{\beta}^T \boldsymbol{X}_i(t)).$$

The accumulated knowledge about what has happened to individuals up to t is denoted by a *history process* $\mathcal{F}_t = \sigma(N_i(s), Y_i(s_+), X_i(t), i = 1, \ldots, n, 0 \leq s \leq t)$.
We assume that the covariates are predictable. We denote the history at an instant just prior to time t by $\mathcal{F}_{t-}$ and have

$$\begin{aligned}\mathsf{E}(dN_i(t)|\mathcal{F}_{t-}) &= Y_i(t)\lambda_i(t)dt \\ &= Y_i(t)\lambda_0(t)\exp(\boldsymbol{\beta}^T \boldsymbol{X}_i(t))dt \\ &= \alpha_i(t)dt.\end{aligned}$$

In the case of right-censored data, the history at time t, $\mathcal{F}_t$, consists of knowledge of pairs (T_i, Δ_i) provided $T_i \leq t$ and the knowledge $T_i > t$ for those individuals still under study at time t. We know (T_i, Δ_i) for those with $T_i < t$ and $T_i \geq t$ for those still under study.
We see that the intensity process α is predictable. The process

$$\mathbf{A}(t) = (A_1(t), \ldots, A_n(t))^T,$$

with

$$A_i(t) = \int_0^t \alpha_i(u)du = \int_0^t Y_i(u)\lambda_i(u)du$$

where $\lambda_i(s) = \lambda_0(s)\exp(\boldsymbol{\beta}^T \boldsymbol{X}_i(t))$, is the compensator of the counting process $\mathbf{N}$.

The covariate values at any time different from a death time are not used in the partial likelihood function. Estimation and testing may proceed as in the Cox model (3.2) with appropriate modifications of $\boldsymbol{X}$ to $\boldsymbol{X}(t)$. This notation allows us to use time-independent covariates as well, for example, if the jth covariate is time-independent, then $\boldsymbol{X}_j(t)$ is constant over time.

Using the counting process approach, the extending partial likelihood function (3.10) can be rewritten as

$$L_n(\boldsymbol{\beta}) = \prod_{j=1}^{n} \left(\frac{\exp(\boldsymbol{\beta}^T \boldsymbol{X}_j(T_j))}{\sum\limits_{i=1}^{n} Y_i(T_j)\exp(\boldsymbol{\beta}^T \boldsymbol{X}_i(T_j))} \right)^{\Delta_j}.$$

The log partial likelihood function based on observation over $[0,\tau]$ is now given by

$$\ell_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{\beta}^T \boldsymbol{X}_i(u) - \log\left\{ \sum_{j=1}^{n} Y_j(u)\exp(\boldsymbol{\beta}^T \boldsymbol{X}_j(u)) \right\} \right] dN_i(u)$$

and the score function is

$$\mathbf{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{X}_i(u) - \frac{\sum\limits_{j=1}^{n} Y_j(u)\boldsymbol{X}_j(u)\exp(\boldsymbol{\beta}^T \boldsymbol{X}_j(u))}{\sum\limits_{j=1}^{n} Y_j(u)\exp(\boldsymbol{\beta}^T \boldsymbol{X}_j(u))} \right] dN_i(u).$$

For covariates which are not time-dependent this is just the score function defined in (3.6).
Andersen and Gill (1982) proved the consistency and asymptotic normality of the MPLE $\widehat{\boldsymbol{\beta}}_n$.
To present their conditions the following functions are defined.

$$S_{0n}(t,\boldsymbol{\beta}) = \frac{1}{n}\sum_{j=1}^{n} Y_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{X}_j(t)), \tag{3.11}$$

$$\mathbf{S}_{1n}(t,\boldsymbol{\beta}) = \frac{1}{n}\sum_{j=1}^{n} Y_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{X}_j(t))\boldsymbol{X}_j(t), \tag{3.12}$$

$$\mathbf{S}_{2n}(t,\boldsymbol{\beta}) = \frac{1}{n}\sum_{j=1}^{n} Y_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{X}_j(t))\boldsymbol{X}_j(t)\boldsymbol{X}_j(t)^T. \tag{3.13}$$

Andersen and Gill (1982) assumed:

A 3.1. *There exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ and scalar, p-vector and $p \times p$ matrix functions $s_0, \mathbf{s}_1$ and $\mathbf{s}_2$, respectively, such that*

a) $\mathbf{S}_{jn}$ *converge in probability (uniformly in $\boldsymbol{\beta} \in \mathcal{B}$ and t) to $\mathbf{s}_j$,* $\quad j = 0, 1, 2;$

b) $\mathbf{s}_j$ *is a continuous function of $\boldsymbol{\beta} \in \mathcal{B}$ uniformly in t and bounded;*

c) s_0 *is bounded away from zero;*

d) $\mathbf{s}_1(t,\boldsymbol{\beta}) = \dfrac{\partial s_0(t,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ *and* $\mathbf{s}_2(t,\boldsymbol{\beta}) = \dfrac{\partial \mathbf{s}_1(t,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$*;* $\boldsymbol{\beta} \in \mathcal{B}$*;*

e) $\mathbf{s}_j(t,\boldsymbol{\beta}_0) = \mathsf{E}\mathbf{S}_{jn}(t,\boldsymbol{\beta}_0), \quad j = 0, 1, 2.$

If **A** 3.1 holds then we define

$$\boldsymbol{\Sigma} = \int_0^{\tau} \mathbf{v}(u,\boldsymbol{\beta}_0)\, s_0(u,\boldsymbol{\beta}_0)\lambda_0(u)du \tag{3.14}$$

with

$$\mathbf{v}(t,\boldsymbol{\beta}_0) = \frac{\mathbf{s}_2(t,\boldsymbol{\beta}_0)}{s_0(t,\boldsymbol{\beta}_0)} - \frac{\mathbf{s}_1(t,\boldsymbol{\beta}_0)\mathbf{s}_1(t,\boldsymbol{\beta}_0)^T}{s_0^2(t,\boldsymbol{\beta}_0)}.$$

A 3.2. *The matrix $\boldsymbol{\Sigma}$ is positive definite.*

The next assumption is a Lindeberg-type condition about the covariates.

A 3.3. *There exists a $\delta > 0$ such that*

$$n^{-1/2}\sup_{i,t} |\boldsymbol{X}_i(t)| Y_i(t)\mathbb{1}(\boldsymbol{\beta}_0^T\boldsymbol{X}_i(t) > -\delta|\boldsymbol{X}_i(t)|) \to 0$$

in probability for $n \to \infty$.

A 3.4. *For* $\tau < \infty$,

$$\int_0^\tau \lambda_0(t)dt < \infty.$$

The observed information matrix has the form

$$\mathbf{I}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \left[\frac{\mathbf{S}_{2n}(u,\boldsymbol{\beta})}{S_{0n}(u,\boldsymbol{\beta})} - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta})\mathbf{S}_{1n}(u,\boldsymbol{\beta})^T}{S_{0n}^2(u,\boldsymbol{\beta})}\right] dN_i(u) \tag{3.15}$$

and can be written as

$$\mathbf{I}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \mathbf{V}_n(u,\boldsymbol{\beta}) dN_i(u) \tag{3.16}$$

with the $p \times p$ matrix

$$\mathbf{V}_n(t,\boldsymbol{\beta}) = \frac{\mathbf{S}_{2n}(t,\boldsymbol{\beta})}{S_{0n}(t,\boldsymbol{\beta})} - \frac{\mathbf{S}_{1n}(t,\boldsymbol{\beta})\mathbf{S}_{1n}(t,\boldsymbol{\beta})^T}{S_{0n}^2(t,\boldsymbol{\beta})}.$$

With assumptions **A** 3.1 and **A** 3.2 the consistency of $\widehat{\boldsymbol{\beta}}_n$ was proved by Andersen and Gill (1982), i.e. for $n \to \infty$

$$\widehat{\boldsymbol{\beta}}_n \xrightarrow{\text{P}} \boldsymbol{\beta}_0.$$

Moreover, under assumptions above, they showed that the score vector is asymptotically normal

$$n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0) \xrightarrow{\text{D}} \mathsf{N}(0,\boldsymbol{\Sigma}),$$

and they also proved the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\text{D}} \mathsf{N}(0,\boldsymbol{\Sigma}^{-1})$$

where $\boldsymbol{\Sigma} = \text{plim}_{n\to\infty} n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0)$.
It follows that $n^{-1}\mathbf{I}_n(\widehat{\boldsymbol{\beta}}_n)$ is a consistent estimator of $\boldsymbol{\Sigma}$ because $\widehat{\boldsymbol{\beta}}_n$ is the consistent estimator.

For later reference it seem to be useful to formulate the basic steps of the verification of the asymptotic normality of $\mathbf{U}_n(\boldsymbol{\beta}_0)$ and then the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$.

The martingale theory implies the decomposition of the counting process into

$$N_i(t) = M_i(t) + A_i(t)$$

where $\mathbf{M}(t) = (M_1(t), ..., M_n(t))^T$ is a n-dimensional martingale.
The martingale decomposition of $dN_i(t)$ then reads

$$\begin{aligned} dN_i(t) &= dM_i(t) + Y_i(t)\lambda_0(t)\exp(\boldsymbol{\beta}_0^T \boldsymbol{X}_i(t))dt \\ &= dM_i(t) + \alpha_i(t)dt. \end{aligned}$$

We obtain for the score function

$$\begin{aligned} &\mathbf{U}_n(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{X}_i(u) - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta})}{S_{0n}(u,\boldsymbol{\beta})} \right] dM_i(u) + \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{X}_i(u) - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta})}{S_{0n}(u,\boldsymbol{\beta})} \right] \alpha_i(u)du. \end{aligned}$$

It is immediately seen that

$$\sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{X}_i(u) - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)}{S_{0n}(u,\boldsymbol{\beta}_0)} \right] Y_i(u)\exp(\boldsymbol{\beta}_0^T \boldsymbol{X}_i(u))\lambda_0(u)du = 0.$$

It follows that the score function evaluated at the true point $\boldsymbol{\beta}_0$ is a (local square integrable) martingale

$$\mathbf{U}_n(\boldsymbol{\beta}_0) = \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{X}_i(u) - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)}{S_{0n}(u,\boldsymbol{\beta}_0)} \right] dM_i(u)$$

where M_i is a zero-mean martingale.
The predictable variation process is denoted $\langle \cdot \rangle$ and the predictable variation process of $n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0)$ is given by

$$\begin{aligned} \mathcal{V}_n(\boldsymbol{\beta}_0) &= \langle n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0) \rangle \\ &= n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{X}_i(u) - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)}{S_{0n}(u,\boldsymbol{\beta}_0)} \right]^{\otimes 2} Y_i(u)\exp(\boldsymbol{\beta}_0^T \boldsymbol{X}_i(u))\lambda_0(u)du \\ &= \int_0^{\tau} \left[\mathbf{S}_{2n}(u,\boldsymbol{\beta}_0) - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)^T}{S_{0n}(u,\boldsymbol{\beta}_0)} \right] \lambda_0(u)du \\ &= \int_0^{\tau} \mathbf{V}_n(u,\boldsymbol{\beta}_0) S_{0n}(u,\boldsymbol{\beta}_0)\lambda_0(u)du. \end{aligned}$$

The assumption formulated by Andersen and Gill (1982) and given above ensure that the Lindeberg-type condition of the martingale central limit theorem is fulfilled so that the distribution of $\mathbf{U}_n(\boldsymbol{\beta}_0)$ converges to a normal distribution with expectation zero and variance matrix $\boldsymbol{\Sigma}$.
Furthermore we have

$$
\begin{aligned}
n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0) = \ & n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\frac{\mathbf{S}_{2n}(u,\boldsymbol{\beta}_0)}{S_{0n}(u,\boldsymbol{\beta}_0)} - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)^T}{S_{0n}^2(u,\boldsymbol{\beta}_0)}\right]dN_i(u) \\
= \ & n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\frac{\mathbf{S}_{2n}(u,\boldsymbol{\beta}_0)}{S_{0n}(u,\boldsymbol{\beta}_0)} - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)^T}{S_{0n}^2(u,\boldsymbol{\beta}_0)}\right]dM_i(u) \\
& + n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\frac{\mathbf{S}_{2n}(u,\boldsymbol{\beta}_0)}{S_{0n}(u,\boldsymbol{\beta}_0)} - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)^T}{S_{0n}^2(u,\boldsymbol{\beta}_0)}\right]\alpha_i(u)du \\
= \ & n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\frac{\mathbf{S}_{2n}(u,\boldsymbol{\beta}_0)}{S_{0n}(u,\boldsymbol{\beta}_0)} - \frac{\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)\mathbf{S}_{1n}(u,\boldsymbol{\beta}_0)^T}{S_{0n}^2(u,\boldsymbol{\beta}_0)}\right]dM_i(u) \\
& + \mathcal{V}_n(\boldsymbol{\beta}_0),
\end{aligned}
$$

that means $\mathcal{V}_n(\boldsymbol{\beta}_0)$ is the compensator of $n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0)$. It follows from the martingale theory that the difference between $\mathcal{V}_n(\boldsymbol{\beta}_0)$ and $n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0)$ converges to zero in probability.
The asymptotic normally of $\widehat{\boldsymbol{\beta}}_n$ follows as indicated in (3.7).

3.4 The Cox model with time-dependent coefficients

In the Cox model (3.2) the coefficients $\boldsymbol{\beta}_0$ are characterized as the risk parameters. Sometimes it seem to be useful that these risk parameters are allowed to change over time. For example, it can happen that the effect of a treatment (not the treatment) vary with time.
This time-dependence is modelled by a further extension of the model. We introduce time-dependent or time-varying coefficients and define the following hazard rate as

$$\lambda(t,\boldsymbol{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T(t)\boldsymbol{x}(t)). \tag{3.17}$$

More general, one can consider the model

$$\lambda(t, \boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T(t)\boldsymbol{x}_1(t) + \boldsymbol{\gamma}^T \boldsymbol{x}_2(t))$$

where $(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t))^T$ is a $(p+q)$-vector of the time-dependent covariate, $\boldsymbol{\beta}(\cdot)$ is a p-dimensional time-dependent coefficients function and $\boldsymbol{\gamma}$ is a q-dimensional constant coefficients. This model has been studied by Martinussen and Scheike (2000), Scheike and Martinussen (2004) and Martinussen and Scheike (2006).

In this thesis, we focus only on the model (3.17) and assume throughout the thesis that the covariates $\boldsymbol{X}$ is time-independent covariates.

The function $\beta_j(\cdot)$ is, as the baseline function λ_0, not specified. Thus, for the estimation of $\beta_j(\cdot)$, nonparametric estimation methods are applied. Let us describe some of them.
Murphy and Sen (1991) proposed the histogram sieve method to estimate the coefficient function $\beta_j(\cdot)$ by assuming that $\beta_j(\cdot)$ is piecewise constant and then obtained an estimator of the cumulative time dependent effects, $B_j(t) = \int_0^t \beta_j(s)ds$. The goodness-of-fit tests based on the sieve estimator was proposed by Murphy (1993) and Marzec and Marzec (1997).
Zucker and Karr (1990) considered the model (3.17) in the case $p = 1$ and assumed that the covariate X does not depend on time. They introduced a penalty functional

$$[\beta, \beta] = \int \beta^{[m]}(s)\beta^{[m]}(s)ds, \quad m \geq s$$

which is the scalar product in the Sobolov space $H^m[0,1]$ of the piecewise m-times differentiable functions β.
As estimator $\widehat{\beta}_n(\cdot)$, referred as penalized partial likelihood estimator (MPPLE), they defined the maximizer of

$$\ell_n(\beta)$$
$$= n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\beta(u)X_i - \log\left\{ \sum_{j=1}^{n} Y_j(u) \exp(\beta(u)X_j) \right\} \right] dN_i(u) - \frac{1}{2}\alpha_n[\beta, \beta].$$

Here $\{\alpha_n\}$ is a sequence of positive numbers, the smoothing parameter, which has to be chosen by the statistician.

The authors showed that under appropriate conditions on β, i.e. $\beta \in H^m[0,1]$, on the convergence of α_n and on the underlying distribution, the MPPLE $\widehat{\beta}_n(t)$ is uniformly consistent with a certain rate of convergence and that is asymptotically normal at an arbitrary fixed point t.
Hastie and Tibshirani (1993) established an algorithm using an iterative strategy to solve the penalized partial likelihood problem.
The smoothing spline estimation in such models were considered by Gray (1992).

Martinussen and Scheike (2000) proposed the cumulative regression coefficients $\widehat{B}_{jn}(t)$ for estimating $B_j(t) = \int_0^t \beta_j(s)ds$. Smoothing $\widehat{B}_{jn}(t)$ by a kernel K and a suitable bandwidth they obtained the desired estimates for β_j. On the basis of these estimators they derived tests for testing the coefficients, these test are of Kolmogorov-Smirnov, Cramer-von Mises type.

Several authors have studied estimators based on local constant or local linear partial likelihood estimators. We will discuss this type of estimators later in Chapter 5.

Let us mention here the following: Local constant or local linear means that locally a constant or a linear function for components $\beta_j(\cdot)$ is fitted.
The parameters are estimated by partial likelihood, and "locally" is expressed by a kernel function and a bandwidth which controls the size of the local neighborhood.
By Taylor's expansion of $\beta_j(s)$ $j = 1, \ldots, p$ in a neighbourhood of t is given by

$$\beta_j(s) \approx \beta_j(t) + \beta_j'(t)(s-t) = b_{1j}(t) + b_{2j}(t)(s-t)$$

where β_j' denotes the first derivative of β_j.
For fixed grid point t, let

$$\boldsymbol{b} = (b_{11}(t), \ldots, b_{1p}(t), b_{21}(t), \ldots, b_{2p}(t))^T$$

and set

$$\tilde{\boldsymbol{X}}_i(u, u-t) = \boldsymbol{X}_i(u) \otimes (1, u-t)^T.$$

Then the log local linear partial likelihood function is given by

$$\ell_t(\boldsymbol{b}) = \sum_{i=1}^{n} \int_0^{\tau} K_{h_n}(u-t) \Bigg[\boldsymbol{b}^T \tilde{\boldsymbol{X}}_i(u, u-t) \\ - \log \Bigg\{ \sum_{j=1}^{n} Y_j(u) \exp(\boldsymbol{b}^T \tilde{\boldsymbol{X}}_j(u, u-t)) \Bigg\} \Bigg] dN_i(u) \tag{3.18}$$

and the local linear partial likelihood estimator of $\boldsymbol{\beta}(\cdot)$ at t is the first p-dimensional component of

$$\widehat{\boldsymbol{b}}(t) = \arg\max_{\boldsymbol{b}} \ell_t(\boldsymbol{b}).$$

Here $K_h(\cdot) = K(\cdot/h)/h$ and K is a kernel function, roughly speaking a density function, $h = h_n$ is the smoothing parameter tending to zero at appropriate rate.

In Chapter 5 we will explain this approach in more detail. Cai and Sun (2003) showed how this estimator can be computed and they established large sample properties as weak pointwise consistency and asymptotic normality at a fixed point t. In the limit theorem the trade-off between bias and variance a common property in nonparametric curve estimation was taken into account. Furthermore the authors proposed a consistent estimator for the asymptotic variance and the cumulative hazard function.
Tests are also considered by Kauermann and Berger (2003). They considered local constant estimators for β_j and as test statistic the partial likelihood ratio is chosen.

Chapter 4

Estimability of the parameter in the Cox model and optimal choice of the covariates

In this chapter the problem of the existence and the uniqueness of the maximum partial likelihood estimator (MPLE) are investigated in more detail. That is, we derive conditions on the covariates which ensure that the score equations (3.5) have a unique solution. As usual, in ML-theory the non-singularity of the Hessian matrix of the partial likelihood function, or in other words of the observed information matrix $\mathbf{I}_n(\boldsymbol{\beta})$, is essential. Therefore, the non-singularity of the observed information matrix will be the basis of the notation of estimability.

Furthermore, we will define the asymptotic estimability. Here we consider sufficient conditions for the non-singularity of the limit of $\mathbf{I}_n(\boldsymbol{\beta})$, i.e. $\boldsymbol{\Sigma}$. Note that the inverse $\boldsymbol{\Sigma}^{-1}$ is the variance matrix in the limit distribution of the estimator $\widehat{\boldsymbol{\beta}}_n$.

The conditions for the asymptotic estimability include assumptions on the underlying distribution of the $\tilde{T}_i$'s and C_i's and on the distribution of the covariates. Also here, our main interest is to investigate the conditions on the covariates.

The study of the influence of the covariates on the estimate can be considered as a problem of experimental design. Experimental design is connected by the derivation of criteria for the optimal choice of the covariates.

Optimal designs under a survival framework were considered in Balakrishnan and Han (2007), Garcet-Rodr´i guez et al. (2008), López-Fidalgo et al.

(2009), Schmidt and Schwabe (2015). In these papers the authors studied the maximum likelihood estimate (MLE). In this thesis we make use of the maximum partial likelihood estimate (MPLE). Consequently, the estimator has different properties and for the experimental design one has to minimize a different criterion. This will be represented in the next sections.

4.1 Estimability of the parameter $\boldsymbol{\beta}_0$

We consider the classical Cox model, which was introduced in Section 3.2. The observations are realizations of i.i.d. $(T_i, \Delta_i, \boldsymbol{X}_i)$, $i = 1, \dots, n$. The true parameter is denoted by $\boldsymbol{\beta}_0$. The parameter $\boldsymbol{\beta}_0$ is estimated by the maximum partial likelihood estimates. Let us recall the partial log likelihood function given at the observed $(t_i, \delta_i, \boldsymbol{x}_i)$, $i = 1, \dots, n$ which is denoted by

$$\ell(\boldsymbol{\beta}; \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x}) = \sum_{i=1}^{n} \delta_i \big(\boldsymbol{\beta}^T \boldsymbol{x}_i - \log n S_{0n}(t_i, \boldsymbol{\beta}) \big)$$

and the observed information matrix

$$\mathbf{I}_n(\boldsymbol{\beta}; \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x}) = \sum_{i=1}^{n} \delta_i \mathbf{V}_n(t_i, \boldsymbol{\beta})$$

where

$$\mathbf{V}_n(t, \boldsymbol{\beta}) = \frac{\mathbf{S}_{2n}(t, \boldsymbol{\beta})}{S_{0n}(t, \boldsymbol{\beta})} - \frac{\mathbf{S}_{1n}(t, \boldsymbol{\beta}) \mathbf{S}_{1n}(t, \boldsymbol{\beta})^T}{S_{0n}^2(t, \boldsymbol{\beta})}.$$

By applying the Cauchy-Schwarz inequality, we obtain that $\mathbf{I}_n(\boldsymbol{\beta}; \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x})$ is non-negative definite for each parameter $\boldsymbol{\beta}$ and all observation $\boldsymbol{t}$, $\boldsymbol{\delta}$ and $\boldsymbol{x}$. Therefore, ℓ is concave. It is strictly concave and has a unique maximum if $\mathbf{I}_n$ is positive definite. Sometimes we use the shorter notation $\mathbf{I}_n(\boldsymbol{\beta})$ instead of $\mathbf{I}_n(\boldsymbol{\beta}; \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x})$. This leads to the following definition:

Definition 4.1. *The parameter $\boldsymbol{\beta}_0$ is called estimable by the maximum partial likelihood estimator at the observed $(\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x})$, if the matrix $\mathbf{I}_n(\boldsymbol{\beta})$ is non-singular for all $\boldsymbol{\beta} \in \mathbb{R}^p$.*

Now, let us derive the sufficient conditions for the non-singularity of $\mathbf{I}_n(\boldsymbol{\beta})$.

Lemma 4.1. *We observe* $(t_i, \delta_i, \boldsymbol{x}_i)$*,* $i = 1, \ldots, n$*. Let* $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m$ *be different support points of the covariates. Then, the observed information matrix can be written as*

$$\mathbf{I}_n(\boldsymbol{\beta}) = \frac{1}{2} \sum_{r=1}^{m} \sum_{s=1}^{m} \kappa_{nrs}(\boldsymbol{\beta}) \mathbf{w}_{rs} \mathbf{w}_{rs}^T \tag{4.1}$$

with

$$\begin{aligned} \mathbf{w}_{rs} &= \boldsymbol{\xi}_r - \boldsymbol{\xi}_s, \\ \kappa_{nrs}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \delta_i \frac{R_r(t_i) R_s(t_i) \exp(\boldsymbol{\beta}^T (\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))}{(\sum_{l=1}^{m} R_l(t_i) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_l))^2} \\ R_l(t) &= \sum_{i=1}^{n} \mathbb{1}(t_i \geq t, \boldsymbol{x}_i = \boldsymbol{\xi}_l). \end{aligned} \tag{4.2}$$

Proof. Consider (3.16), for observations $(\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x})$, we have

$$\mathbf{I}_n(\boldsymbol{\beta}; \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x}) = \sum_{i=1}^{n} \delta_i \mathbf{V}_n(t_i, \boldsymbol{\beta}).$$

The sums S_{jn} introduced in (3.11), (3.12) and (3.13) can be written as

$$\begin{aligned} S_{0n}(t, \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(t_i \geq t) \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i) = \frac{1}{n} \sum_{l=1}^{m} R_l(t) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_l), \\ \mathbf{S}_{1n}(t, \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(t_i \geq t) \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i) \boldsymbol{x}_i = \frac{1}{n} \sum_{l=1}^{m} R_l(t) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_l) \boldsymbol{\xi}_l, \\ \mathbf{S}_{2n}(t, \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(t_i \geq t) \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i) \boldsymbol{x}_i \boldsymbol{x}_i^T = \frac{1}{n} \sum_{l=1}^{m} R_l(t) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_l) \boldsymbol{\xi}_l \boldsymbol{\xi}_l^T. \end{aligned}$$

Obviously, by a simple calculation we have

$$
\begin{aligned}
\mathbf{V}_n(t,\boldsymbol{\beta}) &= \frac{\mathbf{S}_{2n}(t,\boldsymbol{\beta})}{S_{0n}(t,\boldsymbol{\beta})} - \frac{\mathbf{S}_{1n}(t,\boldsymbol{\beta})\mathbf{S}_{1n}(t,\boldsymbol{\beta})^T}{S_{0n}^2(t,\boldsymbol{\beta})} \\
&= \frac{\sum_{r=1}^m \sum_{s=1}^m R_r(t)R_s(t)\boldsymbol{\xi}_r^{\otimes 2}\exp(\boldsymbol{\beta}^T\boldsymbol{\xi}_r + \boldsymbol{\beta}^T\boldsymbol{\xi}_s)}{(\sum_{l=1}^m R_l(t)\exp(\boldsymbol{\beta}^T\boldsymbol{\xi}_l))^2} \\
&\quad - \frac{\sum_{r=1}^m \sum_{s=1}^m R_r(t)R_s(t)\boldsymbol{\xi}_r\boldsymbol{\xi}_s^T\exp(\boldsymbol{\beta}^T\boldsymbol{\xi}_r + \boldsymbol{\beta}^T\boldsymbol{\xi}_s)}{(\sum_{l=1}^m R_l(t)\exp(\boldsymbol{\beta}^T\boldsymbol{\xi}_l))^2} \\
&= \frac{\sum_{r=1}^m \sum_{s=1}^m R_r(t)R_s(t)\boldsymbol{\xi}_r(\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)^T\exp(\boldsymbol{\beta}^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))}{(\sum_{l=1}^m R_l(t)\exp(\boldsymbol{\beta}^T\boldsymbol{\xi}_l))^2} \\
&= \frac{\sum_{r=1}^m \sum_{\substack{s=1 \\ s<r}}^m R_r(t)R_s(t)(\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)^T\exp(\boldsymbol{\beta}^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))}{(\sum_{l=1}^m R_l(t)\exp(\boldsymbol{\beta}^T\boldsymbol{\xi}_l))^2}.
\end{aligned}
$$

Thus, $\mathbf{I}_n(\boldsymbol{\beta})$ can be represented as

$$
\begin{aligned}
\mathbf{I}_n(\boldsymbol{\beta}) &= \sum_{r=1}^m \sum_{\substack{s=1 \\ s<r}}^m (\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)^T \\
&\quad \times \sum_{i=1}^n \delta_i \frac{R_r(t)R_s(t)\exp(\boldsymbol{\beta}^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))}{(\sum_{l=1}^m R_l(t)\exp(\boldsymbol{\beta}^T\boldsymbol{\xi}_l))^2} \\
&= \sum_{r=1}^m \sum_{\substack{s=1 \\ s<r}}^m \mathbf{w}_{rs}\mathbf{w}_{rs}^T\kappa_{nrs}(\boldsymbol{\beta}),
\end{aligned}
\tag{4.3}
$$

where $\mathbf{w}_{rs} = \boldsymbol{\xi}_r - \boldsymbol{\xi}_s$ are p-dimensional vectors characterizing the support points of $\boldsymbol{X}$ and the coefficients

$$
\kappa_{nrs}(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i \frac{R_r(t_i)R_s(t_i)\exp(\boldsymbol{\beta}^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))}{(\sum_{l=1}^m R_l(t_i)\exp(\boldsymbol{\beta}^T\boldsymbol{\xi}_l))^2}
$$

depend on the observations and on the support points of $\boldsymbol{X}$. □

Using relative frequencies instead of $R_l(t_i)$ we obtain

$$f_{il} = \frac{R_l(t_i)}{\sum_{j=1}^m R_j(t_i)} \tag{4.4}$$

and

$$\kappa_{nrs}(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i \frac{f_{ir} f_{is} \exp(\boldsymbol{\beta}^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))}{(\sum_{l=1}^m f_{il} \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_l))^2}. \tag{4.5}$$

Moreover, the coefficients $\kappa_{nrs}(\boldsymbol{\beta})$ are symmetric, i.e., $\kappa_{nrs}(\boldsymbol{\beta}) = \kappa_{nsr}(\boldsymbol{\beta})$ for $s, r = 1, \ldots, m$. We remark that $\kappa_{nrs}(\boldsymbol{\beta})$ depends on the $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m$, too. Therefore, if necessary, we denote these $\kappa_{nrs}(\boldsymbol{\beta})$ by $\kappa_{nrs}(\boldsymbol{\beta}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$. Notice that for the coefficient (4.2) the following holds

$$\begin{aligned}
\sum_{r=1}^m \sum_{s=1}^m \kappa_{nrs}(\boldsymbol{\beta}) &= \sum_{r=1}^m \sum_{s=1}^m \sum_{i=1}^n \delta_i \frac{R_r(t_i) R_s(t_i) \exp(\boldsymbol{\beta}^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))}{(\sum_{l=1}^m R_l(t_i) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_l))^2} \\
&= \sum_{i=1}^n \frac{\delta_i}{(\sum_{l=1}^m R_l(t_i) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_l))^2} \sum_{r=1}^m \sum_{s=1}^m R_r(t_i) R_s(t_i) \exp(\boldsymbol{\beta}^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s)) \\
&= \sum_{i=1}^n \frac{\delta_i}{(\sum_{l=1}^m R_l(t_i) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_l))^2} \sum_{r=1}^m R_r(t_i) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_r) \sum_{s=1}^m R_s(t_i) \exp(\boldsymbol{\beta}^T \boldsymbol{\xi}_s) \\
&= \sum_{i=1}^n \delta_i.
\end{aligned} \tag{4.6}$$

This means that $\sum_{r=1}^m \sum_{s=1}^m \kappa_{nrs}(\boldsymbol{\beta}) = d$ with censoring, and $\sum_{r=1}^m \sum_{s=1}^m \kappa_{nrs}(\boldsymbol{\beta}) = n$ without censoring.

Example 3. Consider $p = 1$ and $m = 2$. For two points $\xi_1, \xi_2 \in \mathbb{R}$ in the support of $\boldsymbol{X} \in \mathbb{R}^1$, we get

$$\mathbf{I}_n(\beta) = \sum_{i=1}^n \delta_i \left[\frac{\sum_{l=1}^2 R_l(t_i) \exp(\beta \xi_l) \xi_l^2}{\sum_{l=1}^2 R_l(t_i) \exp(\beta \xi_l)} - \left\{ \frac{\sum_{l=1}^2 R_l(t_i) \exp(\beta \xi_l) \xi_l}{\sum_{l=1}^2 R_l(t_i) \exp(\beta \xi_l)} \right\}^2 \right]$$

$$
\begin{aligned}
\mathbf{I}_n(\beta) &= \sum_{i=1}^{n} \delta_i \left[\frac{R_1(t_i)\exp(\beta\xi_1)\xi_1^2 + R_2(t_i)\exp(\beta\xi_2)\xi_2^2}{R_1(t_i)\exp(\beta\xi_1) + R_2(t_i)\exp(\beta\xi_2)} \right. \\
&\quad \left. - \left\{ \frac{R_1(t_i)\exp(\beta\xi_1)\xi_1 + R_2(t_i)\exp(\beta\xi_2)\xi_2}{R_1(t_i)\exp(\beta\xi_1) + R_2(t_i)\exp(\beta\xi_2)} \right\}^2 \right] \\
&= \sum_{i=1}^{n} \delta_i \left[\frac{f_{i1}\exp(\beta\xi_1)\xi_1^2 + f_{i2}\exp(\beta\xi_2)\xi_2^2}{f_{i1}\exp(\beta\xi_1) + f_{i2}\exp(\beta\xi_2)} \right. \\
&\quad \left. - \left\{ \frac{f_{i1}\exp(\beta\xi_1)\xi_1 + f_{i2}\exp(\beta\xi_2)\xi_2}{f_{i1}\exp(\beta\xi_1) + f_{i2}\exp(\beta\xi_2)} \right\}^2 \right] \\
&= \sum_{i=1}^{n} \delta_j \left[\Big\{ f_{i1}f_{i2}\exp(\beta\xi_1 + \beta\xi_2)\xi_1^2 - 2f_{i1}f_{j2}\exp(\beta\xi_1 + \beta\xi_2)\xi_1\xi_2 \right. \\
&\quad \left. + f_{j1}f_{j2}\exp(\beta\xi_1 + \beta\xi_2)\xi_2^2 \Big\} \Big/ \Big\{ f_{i1}\exp(\beta\xi_1) + f_{i2}\exp(\beta\xi_2) \Big\}^2 \right] \\
&= \sum_{i=1}^{n} \delta_i \left[\frac{f_{i1}f_{i2}\exp\left(\beta\xi_1 + \beta\xi_2\right)}{\left\{ f_{i1}\exp\left(\beta\xi_1\right) + f_{i2}\exp\left(\beta\xi_2\right) \right\}^2} \right] (\xi_1 - \xi_2)^2.
\end{aligned}
$$

We see that $\mathbf{I}_n(\beta)$ depends on $\mathbf{w}_{12} = \xi_1 - \xi_2$ and the relative frequencies f_{i1}. For positive f_{11}, f_{12} and $\delta_1 = 1$, we have $\mathbf{I}_n(\beta) > 0$ if and only if $\xi_1 \neq \xi_2$. In other words, if we observe lifetimes for two different values of the covariates the observed information matrix is positive (in this case $p = 1$, $m = 2$). Otherwise, if we observe only in one point of the support of $\boldsymbol{X}$ then $\mathbf{I}_n(\boldsymbol{\beta}) = 0$. □

It turns out that we need conditions on the vectors $\mathbf{w}_{rs}$ and on the coefficients in order to ensure that the observed information matrix is positive definite. The condition concerning the support points of $\boldsymbol{X}$ is that the vectors $\mathbf{w}_{rs}$ span the space $\mathbb{R}^p$. However, this is not sufficient, only necessary.

Let $\mathcal{L}(S)$ be the linear space spanned by the elements of the set S. From the representation (4.1) follows that $\mathbf{I}_n(\boldsymbol{\beta})$ is non-singular, if

$$
\mathcal{L}(\{\kappa_{nrs}(\boldsymbol{\beta})\mathbf{w}_{rs}, \quad r = 1, \ldots, m; s = 1, \ldots, m; s < r\}) = \mathbb{R}^p \tag{4.7}
$$

We give now some sufficient conditions for (4.7).

Theorem 4.2. *Under the assumption of Lemma 4.1 we suppose that for some* $i \in \{1, \ldots, n\}$, $\delta_i = 1$ *and* $\prod_{j=1}^{m} f_{ij} > 0$. *Let* $\mathbf{w}_{st} = \boldsymbol{\xi}_s - \boldsymbol{\xi}_t$ *and* $\tilde{m} = \dim \mathcal{L}\{\mathbf{w}_{st} | 1 \leq s < t \leq m\}$.
Then

$$\text{rank}(\mathbf{I}_n(\boldsymbol{\beta})) = \min(p, \tilde{m}).$$

Proof. Recall the form of observed information matrix $\mathbf{I}_n(\boldsymbol{\beta})$:

$$\mathbf{I}_n(\boldsymbol{\beta}) = \frac{1}{2} \sum_{r=1}^{m} \sum_{s=1}^{m} \kappa_{nrs}(\boldsymbol{\beta}) \mathbf{w}_{rs} \mathbf{w}_{rs}^T,$$

where $\mathbf{w}_{rs} = \boldsymbol{\xi}_r - \boldsymbol{\xi}_s$ are a $p \times 1$ vectors, $1 \leq r < s \leq m$ and $\kappa_{nrs}(\boldsymbol{\beta})$ as in (4.5).
The rank of $\mathbf{I}_n(\boldsymbol{\beta})$ is the number of linearly independent vectors in $\{\kappa_{nrs}(\boldsymbol{\beta})\mathbf{w}_{rs}, 1 \leq r < s \leq m\}$.
Because of $f_{ir} > 0$ for all r and $\delta_i = 1$ we have $\kappa_{nrs}(\boldsymbol{\beta}) > 0$ for all r, s, the rank of $\mathbf{I}_n(\boldsymbol{\beta})$ is

$$\text{rank}(\mathbf{I}_n(\boldsymbol{\beta})) = \min(p, \tilde{m}).$$

□

The assumptions of Theorem 4.2 mean that we have at least in one time point uncensored observations for all values $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m$. Under those assumptions, $\boldsymbol{\beta}_0$ is estimable for $\tilde{m} \geq p$. The next theorem gives a slightly more general result.

Theorem 4.3. *If assumption of Lemma 4.1 is fulfilled and if for any* r, s *with* $1 \leq r < s \leq m$ *there exists some* i *with* $\delta_i f_{ir} f_{is} > 0$, *then*

$$\text{rank}(\mathbf{I}_n(\boldsymbol{\beta})) = \min(p, \tilde{m})$$

with $\mathbf{w}_{rs}$ *and* $\tilde{m}$ *as in Theorem 4.2.*

Proof. Using the representation (4.5) for fixed r and s with the condition $\delta_i f_{ir} f_{is} > 0$ for some $i \in \{1, \ldots, n\}$, we get $\kappa_{nrs}(\boldsymbol{\beta}) > 0$. With this condition and (4.1), the desired statement follows. □

A consequence of the Theorem 4.3 is that $m \geq p + 1$ is necessary for the estimability. For $m \leq p$, the parameter $\boldsymbol{\beta}_0$ is not estimable.

4.1.1 Examples

Case 1 : Assume $p = 1$, the hazard function is

$$\lambda(t|x) = \lambda_0(t)\exp(\beta_0 x).$$

If all n observations are taken at one covariate point ξ_1, i.e. $m = 1$, we obtain the log partial likelihood function

$$\begin{aligned}\ell_n(\beta) &= \sum_{i=1}^{n} \delta_i \Big[\beta\xi_1 - \log\Big\{\sum_{j=1}^{n} Y_j(t_i)\exp(\beta\xi_1)\Big\}\Big] \\ &= d\beta\xi_1 - \sum_{i=1}^{n}\delta_i \log\Big\{\exp(\beta\xi_1)R_1(t_i)\Big\}\end{aligned}$$

where $\sum_{i=1}^{n}\delta_i = d$ is the number of uncensored lifetimes. Then, the observed information $\mathbf{I}_n(\beta)$ is

$$\begin{aligned}\mathbf{I}_n(\beta) &= -\frac{\partial^2}{\partial\beta^2}\left[d\beta\xi_1 - \sum_{i=1}^{n}\delta_i\log\Big\{\exp(\beta\xi_1)R_1(t_i)\Big\}\right] \\ &= -\frac{\partial}{\partial\beta}\left[d\xi_1 - \sum_{i=1}^{n}\delta_i\Big\{\frac{\xi_1\exp(\beta\xi_1)R_1(t_i)}{\exp(\beta\xi_1)R_1(t_i)}\Big\}\right] \\ &= 0.\end{aligned}$$

Here, β_0 is not estimable.

Case 2 : Consider $p = 2$ and the hazard function

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}_0^T\boldsymbol{x}).$$

Let $m = 2$, i.e. we choose two distinct points

$$\boldsymbol{\xi}_1 = \begin{pmatrix} z_{11} \\ z_{12} \end{pmatrix}, \quad \boldsymbol{\xi}_2 = \begin{pmatrix} z_{21} \\ z_{22} \end{pmatrix}.$$

Then, the log partial likelihood function is given by

$$\begin{aligned}\ell_n(\boldsymbol{\beta}) = &\sum_{i:\boldsymbol{x}_i=\boldsymbol{\xi}_1} \delta_i(\beta_1 z_{11} + \beta_2 z_{12}) + \sum_{i:\boldsymbol{x}_i=\boldsymbol{\xi}_2} \delta_i(\beta_1 z_{21} + \beta_2 z_{22}) \\ &- \sum_{i=1}^{n}\delta_i\log\Big\{R_1(t_i)\exp(\beta_1 z_{11} + \beta_2 z_{12}) + R_2(t_i)\exp(\beta_1 z_{21} + \beta_2 z_{22})\Big\}\end{aligned}$$

and we can write the observed information matrix $\mathbf{I}_n(\boldsymbol{\beta})$ with relative frequencies as

$$\mathbf{I}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\frac{f_{i1} f_{i2} \exp\left(\beta_1 z_{11} + \beta_2 z_{12} + \beta_1 z_{21} + \beta_2 z_{22}\right)}{\left\{ f_{i1} \exp\left(\beta_1 z_{11} + \beta_2 z_{12}\right) + f_{i2} \exp\left(\beta_1 z_{21} + \beta_2 z_{22}\right) \right\}^2} \right] \times$$
$$\times \begin{pmatrix} z_{11} - z_{21} \\ z_{12} - z_{22} \end{pmatrix}^{\otimes 2}$$

where f_{i1} and f_{i2} as in (4.4) with $m = 2$. Here, the rank $\mathbf{I}_n(\boldsymbol{\beta}) = 1$ if at least for one index i we have $\delta_i f_{i1} f_{i2} > 0$. Otherwise, we have rank $\mathbf{I}_n(\boldsymbol{\beta}) = 0$. In any case, $\boldsymbol{\beta}_0$ is not estimable because rank $\mathbf{I}_n(\boldsymbol{\beta}) < 2 = p$.

Case 3 : We consider the same model as in Case 2, but we choose now three distinct points $m = 3$:

$$\boldsymbol{\xi}_1 = \begin{pmatrix} z_{11} \\ z_{12} \end{pmatrix}, \quad \boldsymbol{\xi}_2 = \begin{pmatrix} z_{21} \\ z_{22} \end{pmatrix}, \quad \boldsymbol{\xi}_3 = \begin{pmatrix} z_{31} \\ z_{32} \end{pmatrix}.$$

With the hazard function

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}_0^T \boldsymbol{x}).$$

the log partial likelihood function is given by

$$\begin{aligned} \ell_n(\boldsymbol{\beta}) = & \sum_{i:\boldsymbol{x}_i=\boldsymbol{\xi}_1} \delta_i(\beta_1 z_{11} + \beta_2 z_{12}) + \sum_{i:\boldsymbol{x}_i=\boldsymbol{\xi}_2} \delta_i(\beta_1 z_{21} + \beta_2 z_{22}) + \sum_{i:\boldsymbol{x}_i=\boldsymbol{\xi}_3} \delta_i(\beta_1 z_{31} + \beta_2 z_{32}) \\ & - \sum_{i=1}^{n} \delta_i \log \Big\{ R_1(t_i) \exp(\beta_1 z_{11} + \beta_2 z_{12}) + R_2(t_i) \exp(\beta_1 z_{21} + \beta_2 z_{22}) \\ & + R_3(t_i) \exp(\beta_1 z_{31} + \beta_2 z_{32}) \Big\} \end{aligned}$$

and the observed information matrix $\mathbf{I}_n(\boldsymbol{\beta})$ with relative frequencies can be written as

$$\mathbf{I}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\frac{\sum_{l=1}^{3} f_{il}(z_{l1}, z_{l2})^T (z_{l1}, z_{l2}) \exp(\beta_1 z_{l1} + \beta_2 z_{l2})}{\sum_{l=1}^{3} f_{il} \exp(\beta_1 z_{l1} + \beta_2 z_{l2})} \right.$$
$$\left. - \left\{ \frac{\sum_{l=1}^{3} f_{il}(z_{l1}, z_{l2})^T \exp(\beta_1 z_{l1} + \beta_2 z_{l2})}{\sum_{l=1}^{3} f_{il} \exp(\beta_1 z_{l1} + \beta_2 z_{l2})} \right\}^{\otimes 2} \right]$$
$$= \sum_{i=1}^{n} \delta_i \left[\sum_{r=1}^{3} \sum_{\substack{s=1 \\ s<r}}^{3} \frac{f_{ir} f_{is} \exp(\beta_1 (z_{r1} + z_{s1}) + \beta_2 (z_{r2} + z_{s2}))}{\{\sum_{l=1}^{3} f_{il} \exp(\beta_1 z_{l1} + \beta_2 z_{l2})\}^2} \times \right.$$
$$\left. \times \begin{pmatrix} z_{r1} - z_{s1} \\ z_{r2} - z_{s2} \end{pmatrix}^{\otimes 2} \right]$$

where f_{il} as in (4.4) with $m = 3$.

From Lemma 4.1 follows rank $\mathbf{I}_n(\boldsymbol{\beta}) = 2$, if $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3$ are linearly independent and $\delta_i f_{ir} f_{is} > 0$ for all r, s and at least one i. Thus, $\boldsymbol{\beta}_0$ is estimable.

Case 4 : Consider $p = 2$ and the hazard function

$$\lambda(t|x) = \lambda_0(t) \exp(\beta_{01} x + \beta_{02} x^2).$$

With $m = 2$ we measure at two distinct points and denote

$$\boldsymbol{\xi}_1 = \begin{pmatrix} z_1 \\ z_1^2 \end{pmatrix}, \quad \boldsymbol{\xi}_2 = \begin{pmatrix} z_2 \\ z_2^2 \end{pmatrix}.$$

So we have a polynomial covariate model.

The log partial likelihood function is given by

$$\ell_n(\boldsymbol{\beta}) = \sum_{i:\boldsymbol{X}_i = \boldsymbol{\xi}_1} \delta_i (\beta_1 z_1 + \beta_2 z_1^2) + \sum_{i:\boldsymbol{X}_i = \boldsymbol{\xi}_2} \delta_i (\beta_1 z_2 + \beta_2 z_2^2)$$
$$- \sum_{i=1}^{n} \delta_i \log \left\{ R_1(t_i) \exp(\beta_1 z_1 + \beta_2 z_1^2) + R_2(t_i) \exp(\beta_1 z_2 + \beta_2 z_2^2) \right\}$$

and we get the observed information matrix $\mathbf{I}_n(\boldsymbol{\beta})$ with relative frequencies as

$$\begin{aligned}\mathbf{I}_n(\boldsymbol{\beta}) =& \sum_{i=1}^{n} \delta_i \Bigg[\frac{f_{i1}\begin{pmatrix} z_1^2 & z_1^3 \\ z_1^3 & z_1^4 \end{pmatrix}\exp(\beta_1 z_1 + \beta_2 z_1^2) + f_{i2}\begin{pmatrix} z_2^2 & z_2^3 \\ z_2^3 & z_2^4 \end{pmatrix}\exp(\beta_1 z_2 + \beta_2 z_2^2)}{f_{i1}\exp(\beta_1 z_1 + \beta_2 z_1^2) + f_{i2}\exp(\beta_1 z_2 + \beta_2 z_2^2)} \\ & - \left\{ \frac{f_{i1}(z_1, z_1^2)^T \exp(\beta_1 z_1 + \beta_2 z_1^2) + f_{i2}(z_2, z_2^2)^T \exp(\beta_1 z_2 + \beta_2 z_2^2)}{f_{i1}\exp(\beta_1 z_1 + \beta_2 z_1^2) + f_{i2}\exp(\beta_1 z_2 + \beta_2 z_2^2)} \right\}^{\otimes 2} \Bigg] \\ =& \sum_{i=1}^{n} \delta_i \left[\frac{f_{i1} f_{i2} \exp\left(\beta_1 z_1 + \beta_2 z_1^2 + \beta_1 z_2 + \beta_2 z_2^2\right)}{\left\{ f_{i1}\exp\left(\beta_1 z_1 + \beta_2 z_1^2\right) + f_{i2}\exp\left(\beta_1 z_2 + \beta_2 z_2^2\right)\right\}^2} \right] \begin{pmatrix} z_1 - z_2 \\ z_1^2 - z_2^2 \end{pmatrix}^{\otimes 2}\end{aligned}$$

where f_{i1} and f_{i2} as in (4.4). Here, the rank $\mathbf{I}_n(\boldsymbol{\beta}) = 1$, if not all $\delta_i f_{i1} f_{i2} = 0$ for $\boldsymbol{\xi}_1 \neq \boldsymbol{\xi}_2$. Otherwise, we have $\mathbf{I}_n(\boldsymbol{\beta}) = 0$. In this polynomial covariate model $\boldsymbol{\beta}_0$ is not estimable.

In the last case, we formulated conditions for the coefficients $\kappa_{nrs}(\boldsymbol{\beta})$ in such a way that we were able to calculate the rank of $\mathbf{I}_n(\boldsymbol{\beta})$. For full rank of the observed information matrix parameter $\boldsymbol{\beta}_0$ is estimable.

4.1.2 Asymptotic estimability

In this section the non-singularity of the limiting matrix of $\mathbf{I}_n(\boldsymbol{\beta})$, i.e. $\boldsymbol{\Sigma} = \operatorname{plim}_{n\to\infty} n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0)$ is considered. We will characterize this by the notation of asymptotic estimability.

Definition 4.2. *The parameter $\boldsymbol{\beta}_0$ is called asymptotically estimable by the maximum partial likelihood estimator if the matrix $\boldsymbol{\Sigma}$ is non-singular for all $\boldsymbol{\beta}_0$.*

To compute the limit $\boldsymbol{\Sigma}$, let us introduce the following (conditional) distribution functions: The conditional distribution function of lifetimes T_i^* is denoted by

$$F(t|\boldsymbol{x}) = \mathsf{P}(\tilde{T} \leq t | \boldsymbol{X} = \boldsymbol{x})$$

the corresponding survival function is, as given in (3.3), $S(t|\boldsymbol{x}) = 1 - F(t|\boldsymbol{x})$. The conditional distribution function of the censoring random variable C_i is denoted by G and we assume that G does not depend on the covariates. Because of the independence assumption, we obtain the conditional distribution of the observations T, which we denote by H:

$$1 - H(t|\boldsymbol{x}) = (1 - F(t|\boldsymbol{x}))(1 - G(t)).$$

As usual, we write $\overline{H} = 1 - H$. By the survival function (3.3) in Cox model, we have

$$\overline{H}(t|\boldsymbol{x}) = \exp(-\Lambda_0(t)\exp(\boldsymbol{\beta}_0^T \boldsymbol{x}))(1 - G(t)). \tag{4.8}$$

Let us assume that

A 4.1. $\mathsf{P}(\boldsymbol{X} = \boldsymbol{\xi}_j) = q_j$ *for* $j = 1, \dots, m$*;* $q_j > 0$*,* $\sum_{j=1}^m q_j = 1$*.*

Theorem 4.4. *Suppose that* **A** *3.1 and* **A** *4.1 are satisfied. Then the matrix* $\boldsymbol{\Sigma}$*, defined in* (3.14) *has the following form*

$$\boldsymbol{\Sigma} = \frac{1}{2}\sum_{r=1}^{m}\sum_{s=1}^{m}\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m)\mathbf{w}_{rs}\mathbf{w}_{rs}^T \tag{4.9}$$

with

$$\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m)$$
$$= \int_0^{\Lambda_0(\tau)} \frac{\exp(-ue^{\boldsymbol{\beta}_0^T\boldsymbol{\xi}_r} + \boldsymbol{\beta}_0^T\boldsymbol{\xi}_r)\exp(-ue^{\boldsymbol{\beta}_0^T\boldsymbol{\xi}_s} + \boldsymbol{\beta}_0^T\boldsymbol{\xi}_s)q_r q_s(1 - G(\Lambda_0^{-1}(u)^-))}{\sum_{j=1}^m \exp(-ue^{\boldsymbol{\beta}_0^T\boldsymbol{\xi}_j} + \boldsymbol{\beta}_0^T\boldsymbol{\xi}_j)q_j} du. \tag{4.10}$$

Proof. We have

$$\boldsymbol{\Sigma} = \int_0^\tau \mathbf{v}(u, \boldsymbol{\beta}_0)\, s_0(u, \boldsymbol{\beta}_0)\lambda_0(u)du$$

with

$$\mathbf{v}(t, \boldsymbol{\beta}_0) = \frac{\mathbf{s}_2(t, \boldsymbol{\beta}_0)}{s_0(t, \boldsymbol{\beta}_0)} - \frac{\mathbf{s}_1(t, \boldsymbol{\beta}_0)\mathbf{s}_1(t, \boldsymbol{\beta}_0)^T}{s_0^2(t, \boldsymbol{\beta}_0)}$$

and

$$\mathbf{s}_j(t, \boldsymbol{\beta}_0) = \mathsf{E}\mathbf{S}_{jn}(t, \boldsymbol{\beta}_0), \qquad j = 0, 1, 2$$

where the expectation is taken with respect to the true underlying distribution.

Let us consider these expectations in more detail. We compute an iterated expectation, first the expectation given the covariates, then we take the expectation with respect to the covariates. We obtain for the conditional expectation

$$\mathsf{E}(Y_i(t)|\boldsymbol{X}_i) = (1 - H(t^-|\boldsymbol{X}_i))$$

and for the expectation

$$\mathsf{E}Y_i(t) = \sum_{r=1}^{m}(1 - H(t^-|\boldsymbol{\xi}_r))q_r.$$

This implies for the functions s_0 and $\mathbf{s}_j$:

$$s_0(t,\boldsymbol{\beta}_0) = \sum_{l=1}^{m}(1 - H(t^-|\boldsymbol{\xi}_l))q_l \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_l),$$

$$\mathbf{s}_1(t,\boldsymbol{\beta}_0) = \sum_{l=1}^{m}(1 - H(t^-|\boldsymbol{\xi}_l))q_l \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_l)\boldsymbol{\xi}_l,$$

$$\mathbf{s}_2(t,\boldsymbol{\beta}_0) = \sum_{l=1}^{m}(1 - H(t^-|\boldsymbol{\xi}_l))q_l \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_l)\boldsymbol{\xi}_l\boldsymbol{\xi}_l^T.$$

and consequently

$$\mathbf{v}(t,\boldsymbol{\beta}_0) = \frac{\sum\limits_{r=1}^{m}\sum\limits_{\substack{s=1\\s<r}}^{m}\overline{H}(t^-|\boldsymbol{\xi}_r)\overline{H}(t^-|\boldsymbol{\xi}_s)}{(\sum\limits_{j=1}^{m}\overline{H}(t^-|\boldsymbol{\xi}_j)q_j\exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_j))^2} q_r q_s \exp(\boldsymbol{\beta}_0^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))\,(\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)^T$$

with

$$\overline{H}(t^-|\boldsymbol{\xi}_j) \neq 0 \quad \text{for} \quad t \in [0,\tau]. \tag{4.11}$$

Thus, $\boldsymbol{\Sigma}$ can be represented as

$$\begin{aligned}\boldsymbol{\Sigma} = &\sum_{r=1}^{m}\sum_{\substack{s=1\\s<r}}^{m}(\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_r - \boldsymbol{\xi}_s)^T q_r q_s \exp(\boldsymbol{\beta}_0^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))\times\\ &\times\int_0^\tau \frac{\overline{H}(u^-|\boldsymbol{\xi}_r)\overline{H}(u^-|\boldsymbol{\xi}_s)}{\sum_{j=1}^{m}\overline{H}(u^-|\boldsymbol{\xi}_j)q_j\exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_j)}\lambda_0(u)du\\ = &\sum_{r=1}^{m}\sum_{\substack{s=1\\s<r}}^{m}\mathbf{w}_{rs}\mathbf{w}_{rs}^T\nu_{rs}(\boldsymbol{\beta}_0,\lambda_0,\mathbf{q},\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_m)\end{aligned}$$

where the (symmetric) coefficients

$$
\begin{aligned}
&\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) \\
&= q_r q_s \exp(\boldsymbol{\beta}_0^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s)) \int_0^\tau \frac{\overline{H}(z^-|\boldsymbol{\xi}_r)\overline{H}(z^-|\boldsymbol{\xi}_s)}{\sum_{j=1}^m \overline{H}(z^-|\boldsymbol{\xi}_j) q_j \exp(\boldsymbol{\beta}_0^T \boldsymbol{\xi}_j)} \lambda_0(z)dz \\
&= \int_0^{\Lambda_0(\tau)} \frac{\exp(-ue^{\boldsymbol{\beta}_0^T\boldsymbol{\xi}_r} + \boldsymbol{\beta}_0^T\boldsymbol{\xi}_r)\exp(-ue^{\boldsymbol{\beta}_0^T\boldsymbol{\xi}_s} + \boldsymbol{\beta}_0^T\boldsymbol{\xi}_s) q_r q_s (1 - G(\Lambda_0^{-1}(u)^-))}{\sum_{j=1}^m \exp(-ue^{\boldsymbol{\beta}_0^T\boldsymbol{\xi}_j} + \boldsymbol{\beta}_0^T\boldsymbol{\xi}_j) q_j} du
\end{aligned}
$$

depend on the underlying baseline distribution, the censoring distribution G, $\boldsymbol{\beta}_0$ and the distribution of covariates. □

There is a connection between the coefficients $\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$ and the $\kappa_{nrs}(\boldsymbol{\beta})$ in (4.5).

Lemma 4.5. *Let the assumptions of Theorem 4.4 be fulfilled. We have*

$$
\sum_{r=1}^m \sum_{s=1}^m \nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) = \int_0^\tau s_0(\boldsymbol{\beta}_0, u)\lambda_0(u)du.
$$

Proof.

$$
\begin{aligned}
&\sum_{r=1}^m \sum_{s=1}^m \nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) = \\
&\qquad = \int_0^\tau \frac{\sum_{r=1}^m \sum_{s=1}^m \overline{H}(u^-|\boldsymbol{\xi}_r)\overline{H}(u^-|\boldsymbol{\xi}_s) q_r q_s \exp(\boldsymbol{\beta}_0^T(\boldsymbol{\xi}_r + \boldsymbol{\xi}_s))}{\sum_{j=1}^m \overline{H}(u^-|\boldsymbol{\xi}_j) q_j \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_j)} \lambda_0(u)du \\
&\qquad = \int_0^\tau \frac{\sum_{r=1}^m \overline{H}(u^-|\boldsymbol{\xi}_r) q_r \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_r) \sum_{s=1}^m \overline{H}(u^-|\boldsymbol{\xi}_s) q_s \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_s)}{\sum_{j=1}^m \overline{H}(u^-|\boldsymbol{\xi}_j) q_j \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_j)} \lambda_0(u)du \\
&\qquad = \int_0^\tau s_0(\boldsymbol{\beta}_0, u)\lambda_0(u)du.
\end{aligned}
$$

□

We have the limit

$$
\boldsymbol{\Sigma} = \operatorname{plim}_{n\to\infty} n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0)
$$

and with the representations (4.9) and (4.1) the limit

$$\frac{1}{n}\kappa_{nrs}(\boldsymbol{\beta}_0, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) \xrightarrow{\mathrm{P}} \nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$$

follows. Moreover, we have

$$\frac{1}{n}\kappa_{nrs}(\widehat{\boldsymbol{\beta}}_n, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) \xrightarrow{\mathrm{P}} \nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$$

and

$$\frac{1}{n}\sum_{r=1}^{m}\sum_{s=1}^{m}\kappa_{nrs}(\boldsymbol{\beta}_0, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) \xrightarrow{\mathrm{P}} \sum_{r=1}^{m}\sum_{s=1}^{m}\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m).$$

This means with (4.6) that $\mathsf{E}\Delta_i$ is equal to $\sum_{r=1}^{m}\sum_{s=1}^{m}\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$.

There are some remarks on the conditions and assumptions:
In any case we have

$$0 \leq \overline{H}(t^-|\boldsymbol{\xi}_r) \leq 1 \quad \text{for all} \quad r.$$

Therefore, we have with (4.10) that

$$\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) \geq 0.$$

In the proof, we assumed the condition (4.11), which is a condition on the distributions F and G and exclude trivial cases, and this condition is fulfilled in general.
Because of

$$0 < \int_0^\tau \lambda_0(u)du < \infty$$

and (4.8), we obtain (4.11). Hence, (4.11) is a condition on the choice of τ and the parameter space.
Furthermore, we will exclude that censoring arises only in intervals where no observation is possible. Therefore, we assume

$$\int_0^\tau (1 - G(u^-))\lambda_0(u)du > 0. \tag{4.12}$$

This condition ensures that $\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) > 0$.

Theorem 4.6. *We consider the Cox model and suppose that the assumptions of Theorem 4.4 are fulfilled. Then, $\boldsymbol{\Sigma}$ is positive definite if and only if*

$$\dim \mathcal{L}(\{\boldsymbol{\xi}_r - \boldsymbol{\xi}_s, 1 \leq s < r \leq m\}) = p. \tag{4.13}$$

Proof. We have

$$1 - F(t|\boldsymbol{\xi}_l) = \exp\Big(- \int_0^t \lambda_0(u)du \exp(\boldsymbol{\beta}_0^T \boldsymbol{\xi}_l)\Big)$$

and together with (4.12) $\nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m) > 0$ for all $1 \leq s < r \leq m$ follows. Then, it is clear that $\operatorname{rank} \boldsymbol{\Sigma} = \dim \mathcal{L}(\{\mathbf{w}_{rs}, 1 \leq s < r \leq m\})$. Thus, $\boldsymbol{\beta}_0$ is asymptotically estimable if and only if (4.13) is fulfilled. □

The importance of this result is that the asymptotic estimability depends only on the support points of $\boldsymbol{X}$. Here, we remark that (4.12) is a condition, which is fulfilled in general, only in unimportant trivial cases (4.12) does not hold.

4.1.3 Asymptotic variance for general covariates

In Theorem 4.4, a representation of the inverse of the asymptotic variance is given under the assumption that the support of $\boldsymbol{X}$ is finite. Now we will find a representation of $\boldsymbol{\Sigma}$ for a general distribution of the covariates $\boldsymbol{X}$. Under **A** 4.1, the support points are $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m$ and we have

$$\mathsf{P}(\boldsymbol{X} = \boldsymbol{\xi}_j) = q_j \quad \text{for} \quad j = 1, \ldots, m$$

in which $q_j > 0$, $\sum_{j=1}^m q_j = 1$ and the corresponding induced measure Q in $\mathbb{R}^p$ is defined by

$$Q(\boldsymbol{\zeta}) = \begin{cases} q_j, & \text{if } \boldsymbol{\zeta} = \boldsymbol{\xi}_j \\ 0, & \text{otherwise .} \end{cases}$$

The matrix $\boldsymbol{\Sigma}$ in 4.9 depends on $\boldsymbol{\beta}_0, \lambda_0, G, Q$ and therefore we use the notation $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$.
We have the representations

$$\int_0^\tau \overline{H}(t^-|\boldsymbol{\zeta}) \exp(\boldsymbol{\beta}_0^T \boldsymbol{\zeta}) dQ(\boldsymbol{\zeta}) = \sum_{j=1}^m \overline{H}(t^-|\boldsymbol{\xi}_j) q_j \exp(\boldsymbol{\beta}_0^T \boldsymbol{\xi}_j),$$

and

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q) = \frac{1}{2}\int\int h(\boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\beta}_0, \lambda_0, G, Q)(\boldsymbol{\zeta}-\boldsymbol{\rho})(\boldsymbol{\zeta}-\boldsymbol{\rho})^T dQ(\boldsymbol{\zeta})dQ(\boldsymbol{\rho}) \tag{4.14}$$

with

$$h(\boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\beta}_0, \lambda_0, G, Q) = \int_0^\tau \frac{\overline{H}(u^-|\boldsymbol{\zeta})\exp(\boldsymbol{\beta}_0^T\boldsymbol{\zeta})\overline{H}(u^-|\boldsymbol{\rho})\exp(\boldsymbol{\beta}_0^T\boldsymbol{\rho})}{\int \overline{H}(u^-|\boldsymbol{\eta})\exp(\boldsymbol{\beta}_0^T\boldsymbol{\eta})dQ(\boldsymbol{\eta})}\lambda_0(u)du. \tag{4.15}$$

and

$$\begin{aligned}\int\int h(\boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\beta}_0, \lambda_0, G, Q)dQ(\boldsymbol{\zeta})dQ(\boldsymbol{\rho}) &= \sum_{r=1}^m\sum_{s=1}^m q_r q_s h(\boldsymbol{\xi}_r, \boldsymbol{\xi}_s, \boldsymbol{\beta}_0, \lambda_0, G, Q)\\ &= \sum_{r=1}^m\sum_{s=1}^m \nu_{rs}(\boldsymbol{\beta}_0, \lambda_0, G, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m).\end{aligned}$$

Up to now Q is a measure with a finite support. For fixed $\boldsymbol{\beta}_0, \lambda_0, Q$, the matrix $\boldsymbol{\Sigma}$ is a continuous functional on Q. Therefore, $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ from (4.14) is defined by continuous continuation for all probability measures, which are limits of measures with a finite support.

Definition 4.3. *Let $\boldsymbol{\mathcal{X}} \subseteq \mathbb{R}^p$ and let $\mathfrak{B}(\boldsymbol{\mathcal{X}})$ be the Barel σ-algebra of $\boldsymbol{\mathcal{X}}$. Assume that the assumptions* **A** *3.1,* **A** *3.2 and* **A** *3.3 are satisfied. For a probability measure Q over $(\boldsymbol{\mathcal{X}}, \mathfrak{B}(\boldsymbol{\mathcal{X}}))$ and with the representations* (4.15) *and*

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q) = \frac{1}{2}\int\int h(\boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\beta}_0, \lambda_0, G, Q)(\boldsymbol{\zeta}-\boldsymbol{\rho})(\boldsymbol{\zeta}-\boldsymbol{\rho})^T dQ(\boldsymbol{\zeta})dQ(\boldsymbol{\rho}),$$

the matrix $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ is called the asymptotic variance matrix of the MPLE of $\boldsymbol{\beta}_0$ in the model (3.2) *where the covariates $\boldsymbol{X}$ have the distribution Q.*

We mention some properties of $\boldsymbol{\Sigma}$. At first the influence of the baseline hazard rate is discussed and then we state the continuity of $\boldsymbol{\Sigma}$.

Theorem 4.7. *Let Q be the induced measure of the covariates $\boldsymbol{X}$. The matrix $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ depends on the baseline hazard rate only via the cumulative baseline hazard rate Λ_0.Then we have the representation (4.14) with*

$$h(\boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\beta}_0, \lambda_0, G, Q) = \int_0^{\Lambda_0(\tau)} \frac{\exp\Big(- ue^{\boldsymbol{\beta}_0^T \boldsymbol{\zeta}} - ue^{\boldsymbol{\beta}_0^T \boldsymbol{\rho}} + \boldsymbol{\beta}_0^T(\boldsymbol{\zeta} + \boldsymbol{\rho})\Big)\Big(1 - G(\Lambda_0^{-1}(u)^-)\Big)}{\int \exp\Big(- ue^{\boldsymbol{\beta}_0^T \boldsymbol{\eta}} + \boldsymbol{\beta}_0^T \boldsymbol{\eta}\Big) dQ(\boldsymbol{\eta})} du. \tag{4.16}$$

Proof. We substitute (4.8) in (4.15) and obtain

$$h(\boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\beta}_0, \lambda_0, G, Q) = \int_0^{\tau} \frac{\exp\big(- \Lambda_0(z^-)e^{\boldsymbol{\beta}_0^T \boldsymbol{\zeta}}\big) e^{\boldsymbol{\beta}_0^T \boldsymbol{\zeta}} \exp\big(- \Lambda_0(z^-)e^{\boldsymbol{\beta}_0^T \boldsymbol{\rho}}\big) e^{\boldsymbol{\beta}_0^T \boldsymbol{\rho}} (1 - G(z^-))^2}{\int \exp\big(- \Lambda_0(z^-)e^{\boldsymbol{\beta}_0^T \boldsymbol{\eta}}\big) e^{\boldsymbol{\beta}_0^T \boldsymbol{\eta}} (1 - G(z^-)) dQ(\boldsymbol{\eta})} \lambda_0(z) dz.$$

Hence we have

$$h(\boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\beta}_0, \lambda_0, G, Q) = \int_0^{\tau} \frac{\exp\big(- \Lambda_0(z^-)e^{\boldsymbol{\beta}_0^T \boldsymbol{\zeta}} + \boldsymbol{\beta}_0^T \boldsymbol{\zeta}\big) \exp\big(- \Lambda_0(z^-)e^{\boldsymbol{\beta}_0^T \boldsymbol{\rho}} + \boldsymbol{\beta}_0^T \boldsymbol{\rho}\big)(1 - G(z^-))}{\int \exp\Big(- \Lambda_0(z^-)e^{\boldsymbol{\beta}_0^T \boldsymbol{\eta}} + \boldsymbol{\beta}_0^T \boldsymbol{\eta}\Big) dQ(\boldsymbol{\eta})} d\Lambda_0(z)$$

and with an integral transformation we get

$$h(\boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\beta}_0, \lambda_0, G, Q) = \int_0^{\Lambda_0(\tau)} \frac{\exp\Big(- ue^{\boldsymbol{\beta}_0^T \boldsymbol{\zeta}} + \boldsymbol{\beta}_0^T \boldsymbol{\zeta}\Big) \exp\Big(- ue^{\boldsymbol{\beta}_0^T \boldsymbol{\rho}} + \boldsymbol{\beta}_0^T \boldsymbol{\rho}\Big)\Big(1 - G(\Lambda_0^{-1}(u)^-\Big)}{\int \exp(-ue^{\boldsymbol{\beta}_0^T \boldsymbol{\eta}} + \boldsymbol{\beta}_0^T \boldsymbol{\eta}) dQ(\boldsymbol{\eta})} du.$$

This is the desired representation. □

With the asymptotic variance matrix $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ we are able to characterize the influence of a covariate $\boldsymbol{X}$ with the induced measure Q. Moreover, we can compare two measures Q_1 and Q_2 by comparing $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_0, \lambda_0, G, Q_1)$ with $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_0, \lambda_0, G, Q_2)$. This is the basis for finding optimal covariates. An important property is the continuity of $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ or of their inverse $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$. For the parameters in Euclidean spaces the usual Euclidean norms are taken. By assumption λ_0 is continuous, i.e. $\lambda_0 \in C([0,\tau])$ for the space $C([0,\tau])$ of continuous functions over the interval $[0,\tau]$ and the sup norm

$$\|\lambda_0\| = \sup_{t\in[0,\tau]} |\lambda_0(t)|$$

will be used. In the space of measures over $(\mathcal{X}, \mathfrak{B}(\mathcal{X}))$ we use as the distance the total variation distance

$$\|Q_1 - Q_2\| = \sup_{B\in\mathfrak{B}(\mathcal{X})} |Q_1(B) - Q_2(B)|$$

for two measures Q_1, Q_2.

Lemma 4.8. *Assume that $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ and $\lambda_0 \in C([0,\tau])$. Let G be a censoring distribution and let Q be a measure over $\mathfrak{B}(\mathcal{X})$. Then, $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ is a continuous function in their arguments.*

Proof. We start with the representation of h from (4.16), so we can see that this h is a continuous function in all arguments. Using the representations (4.14), it can be show that $\boldsymbol{\Sigma}$ is continuous in $\boldsymbol{\beta}_0, \lambda_0, G, Q$. □

4.2 Optimal covariates

Now, let us assume that the limit $\boldsymbol{\Sigma}$ of the observation matrix is non-singular. As already shown, $\boldsymbol{\Sigma}$ depends on the underlying conditional distribution F via the baseline hazard rate λ_0 and the parameter $\boldsymbol{\beta}_0$, on the censoring distribution G and on the distribution of the covariates. In this section, we suppose that the underlying distribution F and the distribution G are determined by the process under the study. However, we assume that we have some freedom for choosing the covariates.
The aim is to choose the covariates in such a way that the asymptotic variance of the MPLE $\widehat{\boldsymbol{\beta}}_n$ is in some sense "small". To do this, we consider the covariate values $\boldsymbol{x}$ as before as the realizations of a random variable with a distribution

Q. Choosing the covariates in an optimal way means to determine an optimal distribution or optimal probability measure Q^*. Roughly speaking, this task is a problem of experimental design. Classical experimental design is based on the information matrix or variance matrix; one minimizes some functional of these matrices to find an optimal design. These functionals are convex functionals of Q.
For such situations, the optimal designs can be characterized by general expressions. Results of Whittle (1973) solve the optimization problem. Equivalence theorems and iteration procedures due to Kiefer (1961), Fedorov (1972), Wynn (1972), Läuter and Läuter (1984) and López-Fidalgo et al. (2009). Here we use similar principles. We know $\mathbf{\Sigma}$ and its dependence on Q. Unfortunately, we could not prove the convexity of functionals. However, based on the results of asymptotic estimability the calculation of suboptimal covariates is considered. For special cases suboptimal covariates are to be found.

4.2.1 Local optimal covariates

We will see that the optimal measure Q^* depends on the underlying F, i.e. on $\boldsymbol{\beta}_0$ and λ_0, and on G. Therefore, we define the notation of local D-optimality. To emphasize the dependence of $\mathbf{\Sigma}$ on $\boldsymbol{\beta}_0$, λ_0 and Q, let us write $\mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$.

Definition 4.4. *Let $\mathcal{D}$ be a set of probability measures on $\mathbb{R}^p$. The measure $Q^* \in \mathcal{D}$ is called D-optimal in $\mathcal{D}$ at fixed $\boldsymbol{\beta}_0$, λ_0 and G if*

$$\det \mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q^*) = \max_{Q \in \mathcal{D}} \det \mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q).$$

The random variable $\boldsymbol{X}^*$, which has the induced measure Q^*, is called a locally D-optimal covariate. We suppress in Q^* the dependence on G.

The set $\mathcal{D}$ of measures includes the choice of the support of the covariates. This choice depends strongly on the practical background of the model. We will choose as a standard set for the possible support points the p-dimensional unit cube $[0,1]^p$. The probability measure can be a discrete measure or any other measure on $[0,1]^p$ with the corresponding Borel-σ-algebra.

In the next example, the explicit representation of $\mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ for two-points covariates is calculated and the local optimal two-points covariates are given.

Example 4. Consider model (3.2) with $p = 1$ and $x \in [0,1]$. It is known that at least $m = 2$ different values of covariates are necessary to estimate β_0. Let us determine a local D-optimal two-points covariate X. As a submodel we choose an exponential model, i.e. the baseline hazard is constant,

$$\lambda(t|x) = \lambda_0 \exp(\beta_0 x), \tag{4.17}$$

$\beta_0, x \in \mathbb{R}^1$ and let be no censoring up to $\tau > 0$. We put

$$G(t) = 1 - \delta_{[0,\tau)}(t) \quad \text{for} \quad t \geq 0,$$

where $\delta_{[0,\tau)}(t)$ is the indicator function of $[0,\tau)$, i.e.

$$\delta_{[0,\tau)}(t) = \begin{cases} 1 & \text{if } t \in [0,\tau), \\ 0 & \text{otherwise.} \end{cases}$$

Let X be the two-points covariate with

$$\mathsf{P}(X = \xi_j) = q_j \quad \text{for} \quad j = 1,2, \quad \xi_1, \xi_2 \in [0,1]$$

and $q_1 + q_2 = 1$.
The measure Q is defined by

$$Q(\eta) = \begin{cases} q_j, & \text{if } \eta = \xi_j \quad j = 1,2, \\ 0, & \text{otherwise.} \end{cases}$$

From (4.14), we get

$$\mathbf{\Sigma}(\beta_0, \lambda_0, G, Q) = \frac{1}{2} \int \int h(\zeta, \rho, \beta_0, \lambda_0, G, Q)(\zeta - \rho)^2 dQ(\zeta) dQ(\rho)$$

with

$$h(\zeta, \rho, \beta_0, \lambda_0, G, Q) = \int_0^\tau \frac{\overline{H}(u^-|\zeta)\exp(\beta_0\zeta)\overline{H}(u^-|\rho)\exp(\beta_0\rho)}{\int \overline{H}(u^-|\eta)\exp(\beta_0\eta)dQ(\eta)} \lambda_0(u) du.$$

Under the above mentioned censoring distribution, we have

$$1 - H(t|\xi) = \exp(-\lambda_0 \exp(\beta_0 \xi) t) \quad \text{for} \quad t \in [0,\tau) \tag{4.18}$$

and $1 - H(t|\xi) = 0$ for $t \geq \tau$.

Hence, we obtain with (4.18)

$$\int \overline{H}(t^-|\eta)\exp(\beta_0\eta)dQ(\eta)$$
$$= q_1\exp(-\lambda_0\exp(\beta_0\xi_1)t+\beta_0\xi_1)+q_2\exp(-\lambda_0\exp(\beta_0\xi_2)t+\beta_0\xi_2).$$

For the calculation of $\mathbf{\Sigma}(\beta_0,\lambda_0,G,Q)$, we need moreover the following representations, so we put

$$g_j(t,\beta_0,\lambda_0) := q_j\exp(-\lambda_0\exp(\beta_0\xi_j)t+\beta_0\xi_j), \quad j=1,2.$$

To formulate $\mathbf{\Sigma}$, we have

$$\begin{aligned}
&\iint \overline{H}(t^-|\zeta)e^{(\beta_0\zeta)}\overline{H}(t^-|\rho)e^{(\beta_0\rho)}(\zeta-\rho)^2dQ(\zeta)dQ(\rho)\\
&=\iint e^{(-\lambda_0\exp(\beta_0\zeta)t+\beta_0\zeta)}e^{(-\lambda_0\exp(\beta_0\rho)t+\beta_0\rho)}(\zeta-\rho)^2dQ(\zeta)dQ(\rho)\\
&=\int\Big\{q_1e^{(-\lambda_0\exp(\beta_0\xi_1)t+\beta_0\xi_1)}(\xi_1-\rho)^2+q_2e^{(-\lambda_0\exp(\beta_0\xi_2)t+\beta_0\xi_2)}(\xi_2-\rho)^2\Big\}\times\\
&\quad\times\overline{H}(t^-|\rho)e^{(\beta_0\rho)}dQ(\rho)\\
&=\int\Big\{g_1(t,\beta_0,\lambda_0)(\xi_1-\rho)^2+g_2(t,\beta_0,\lambda_0)(\xi_2-\rho)^2\Big\}\overline{H}(t^-|\rho)e^{(\beta_0\rho)}dQ(\rho)\\
&=g_2(t,\beta_0,\lambda_0)(\xi_2-\xi_1)^2g_1(t,\beta_0,\lambda_0)+g_1(t,\beta_0,\lambda_0)(\xi_1-\xi_2)^2g_2(t,\beta_0,\lambda_0)\\
&=2g_1(t,\beta_0,\lambda_0)g_2(t,\beta_0,\lambda_0)(\xi_2-\xi_1)^2.
\end{aligned}$$

The representation of $\mathbf{\Sigma}$ is given by

$$\begin{aligned}
&\mathbf{\Sigma}(\beta_0,\lambda_0,G,Q)\\
&=q_1q_2(\xi_1-\xi_2)^2\exp(\beta_0\xi_1+\beta_0\xi_2)\lambda_0\times\\
&\quad\times\int_0^\tau\frac{\exp(-\lambda_0\exp(\beta_0\xi_1)u-\lambda_0\exp(\beta_0\xi_2)u)}{q_1\exp(-\lambda_0\exp(\beta_0\xi_1)u+\beta_0\xi_1)+q_2\exp(-\lambda_0\exp(\beta_0\xi_2)u+\beta_0\xi_2)}du\\
&=q_1q_2(\xi_1-\xi_2)^2\exp(\beta_0(\xi_1+\xi_2))\lambda_0\times\\
&\quad\times\int_0^\tau\frac{1}{q_1\exp(\lambda_0\exp(\beta_0\xi_2)u+\beta_0\xi_1)+q_2\exp(\lambda_0\exp(\beta_0\xi_1)u+\beta_0\xi_2)}du.
\end{aligned}$$

For obtaining local optimal two-points covariate we have to maximize $\mathbf{\Sigma}(\beta_0,\lambda_0,G,Q)$ with respect to q_1,ξ_1,q_2,ξ_2 under the restrictions

$$q_1,q_2\geq 0, \quad q_1+q_2=1,\xi_1,\xi_2\in[0,1].$$

This can be done numerically. Results of a computation for several values of $\boldsymbol{\beta}_0$ and λ_0 are given in Table 4.1.
Using equation (4.9) and (4.10) with $m = 2$, the scalar $\boldsymbol{\Sigma}$ is maximized w.r.t. Q. Denote

$$\max_Q \boldsymbol{\Sigma}(\beta_0, \lambda_0, G, Q) =: \boldsymbol{\Sigma}^*(\beta_0, \lambda_0, G).$$

One sees that the local optimal covariates depend on the chosen parameters. Especially, for $\lambda_0 = 1$, one sees that for $\beta_0 = 1.0$ the largest value of $\boldsymbol{\Sigma}(1, 1, G, Q)$ is reached for $\xi_1^* = 0, \xi_2^* = 1$. For $\beta_0 = 6.0$ we see that $\boldsymbol{\Sigma}(6, 1, G, Q)$ is maximal for $\xi_1^* = 0.4$, $\xi_2^* = 1$. We find out that the end points of the interval $[0, 1]$ are not included necessarily in the local optimal covariate. This depends on the value of β_0. As visible in Figure A.1, for small β_0 we expect that the end points are local optimal covariate points, for larger values (e.g. $\beta_0 = 6$) this will not be the case.

Table 4.1: Local optimal two-points covariates of the exponential model (4.17) as a submodel of (3.2).

	$\lambda_0 = 1$				$\lambda_0 = 3$			
β_0	ξ_1^*	ξ_2^*	q_1^*	$\boldsymbol{\Sigma}^*(\beta_0, \lambda_0, G)$	ξ_1^*	ξ_2^*	q_1^*	$\boldsymbol{\Sigma}^*(\beta_0, \lambda_0, G)$
1	0	1	0.52431	0.20525	0	1	0.52430	0.20525
1.5	0	1	0.55200	0.16915	0	1	0.55195	0.16915
2	0	1	0.58355	0.13400	0	1	0.58365	0.13400
2.5	0	1	0.61605	0.10285	0	1	0.61490	0.10280
3	0	1	0.64760	0.07680	0	1	0.65595	0.07700
4	0.1	0.9	0.65730	0.04350	0	0.8	0.66135	0.04415
5	0.2	0.9	0.66075	0.02825	0.3	1	0.6406	0.03675
6	0.4	1	0.67585	0.02500	0.3	0.8	0.63630	0.02435
7	0.4	0.9	0.57845	0.01840	0.2	0.7	0.66580	0.01845
8	0.3	0.8	0.59780	0.01355	0.2	0.6	0.65710	0.01390

In Figure A.2 the similar plots are represented with $\lambda_0 = 3$ and the dependence of $\boldsymbol{\Sigma}(\beta_0, \lambda_0, G, Q)$ and the ξ_1, ξ_2 is visible. Qualitatively, there is no difference to the behavior under $\lambda_0 = 1$. In Figure A.3 the dependence between $\boldsymbol{\Sigma}(\beta_0, \lambda_0, G, Q)$ and the weight q_1 is plotted. These plots give some hints for the fact that local optimal covariates can correspond to a weight $q_1 \neq 0.5$. In addition, we see that $\boldsymbol{\Sigma}(\beta_0, \lambda_0, G, Q)$ is a concave function of q_1. This appears in special cases the concavity of $\boldsymbol{\Sigma}$ as function of q_1. In general we cannot prove this concavity.

□

For finding local optimal covariates in classes of m-point measures Q with the support points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m$ and

$$\mathsf{P}(\boldsymbol{X} = \boldsymbol{\xi}_j) = q_j, \quad q_j \geq 0, \quad \text{for} \quad j = 1, \ldots, m \tag{4.19}$$

$$\sum_{j=1}^{m} q_j = 1 \tag{4.20}$$

one needs the form of $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ too. In Lemma 4.9 the $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ is calculated for the special $G(t) = 1 - \delta_{[0,\tau)}(t)$.

Lemma 4.9. *Consider the Cox model*

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}_0^T \mathbf{x})$$

without censoring up to τ*. We assume* G *as in Example 4. Let the support of* $\boldsymbol{X}$ *be finite with* (4.19) *and* (4.20)*. We use* Λ_0 *as the cumulative hazard rate,*

$$\Lambda_0(t) = \int_0^t \lambda_0(\zeta) d\zeta.$$

Then we have

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q) = \frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{m} q_i q_j \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_i)\exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_j)(\boldsymbol{\xi}_i - \boldsymbol{\xi}_j)(\boldsymbol{\xi}_i - \boldsymbol{\xi}_j)^T$$
$$\times \int_0^\tau \frac{\exp(-\Lambda_0(u)\exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_i) - \Lambda_0(u)\exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_j))}{\sum_{l=1}^m q_l \exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_l)\exp(-\Lambda_0(u)\exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_l))}\lambda_0(u)du. \tag{4.21}$$

Proof. To formulate representation $\boldsymbol{\Sigma}$, we substitute

$$g_i(t, \boldsymbol{\beta}_0, \lambda_0) := q_i \exp(-\Lambda_0(t)\exp(\boldsymbol{\beta}_0^T\boldsymbol{\xi}_i) + \boldsymbol{\beta}_0^T\boldsymbol{\xi}_i). \tag{4.22}$$

Then we have

$$\overline{H}(t^-|\boldsymbol{x})\exp(\boldsymbol{\beta}_0^T\boldsymbol{x}) = \exp(-\Lambda_0(t)\exp(\boldsymbol{\beta}_0^T\boldsymbol{x}) + \boldsymbol{\beta}_0^T\boldsymbol{x}).$$

With similar calculations as in Example 4 we obtain

$$\int \overline{H}(t^-|\boldsymbol{\eta}) \exp(\boldsymbol{\beta}_0^T \boldsymbol{\eta}) dQ(\boldsymbol{\eta}) = \sum_{i=1}^{m} g_i(t, \boldsymbol{\beta}_0, \lambda_0),$$

$$\int\int \overline{H}(t^-|\boldsymbol{\zeta}) \exp(\boldsymbol{\beta}_0^T \boldsymbol{\zeta}) \overline{H}(t^-|\boldsymbol{\rho}) \exp(\boldsymbol{\beta}_0^T \boldsymbol{\rho}) \lambda_0(t) \times$$
$$\times (\boldsymbol{\zeta}-\boldsymbol{\rho})(\boldsymbol{\zeta}-\boldsymbol{\rho})^T dQ(\boldsymbol{\zeta}) dQ(\boldsymbol{\rho}) = \sum_{j=1}^{m}\sum_{i=1}^{m} g_j(t, \boldsymbol{\beta}_0, \lambda_0) g_i(t, \boldsymbol{\beta}_0, \lambda_0)(\boldsymbol{\xi}_i - \boldsymbol{\xi}_j)(\boldsymbol{\xi}_i - \boldsymbol{\xi}_j)^T.$$

Hence, the desired representation for $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ follows. □

The result of this Lemma 4.9 can be extended for general covariates with any measure Q. It is easy to calculate with (4.14), that is

$$\boldsymbol{\Sigma}(\beta_0, \lambda_0, G, Q) = \frac{1}{2} \int\int h(\zeta, \rho, \beta_0, \lambda_0, G, Q)(\zeta - \rho)(\zeta - \rho)^T dQ(\zeta) dQ(\rho)$$

with

$$\begin{aligned} & h(\zeta, \rho, \beta_0, \lambda_0, G, Q) \\ &= \int_0^{\tau} \frac{\exp(-\Lambda_0(u)\exp(\boldsymbol{\beta}_0^T\zeta) + \boldsymbol{\beta}_0^T\zeta)\exp(-\Lambda_0(u)\exp(\boldsymbol{\beta}_0^T\rho) + \boldsymbol{\beta}_0^T\rho)}{\int \exp(-\Lambda_0(u)\exp(\boldsymbol{\beta}_0^T\eta) + \boldsymbol{\beta}_0^T\eta) dQ(\eta)} d\Lambda_0(u) \\ &= \int_0^{\Lambda_0(\tau)} \frac{\exp(-u\exp(\boldsymbol{\beta}_0^T\zeta) + \boldsymbol{\beta}_0^T\zeta)\exp(-u\exp(\boldsymbol{\beta}_0^T\rho) + \boldsymbol{\beta}_0^T\rho)}{\int \exp(-u\exp(\boldsymbol{\beta}_0^T\eta) + \boldsymbol{\beta}_0^T\eta) dQ(\eta)} du. \end{aligned}$$

In this representation, it is obvious that the baseline hazard rate has an influence to $\boldsymbol{\Sigma}$ only via the fixed $\Lambda_0(\tau)$ in the upper boundary in the integral of $h(\zeta, \rho, \beta_0, \lambda_0, G, Q)$. That is, the reason why the value of $\boldsymbol{\Sigma}$ will not strongly change for different λ_0.
Numerically, it is not a problem to calculate local optimal covariates for other submodels. The next example considers a Weibull model.

Example 5. Consider model (3.2) with $p = 1$ and assume the baseline is a Weibull distribution with parameters $\theta > 0$ and $\mu > 0$

$$\lambda_0(t|x) = \theta\mu t^{\mu-1} \exp(\beta_0 x), \tag{4.23}$$

$\beta_0, x \in \mathbb{R}^1$. The censoring distribution G is the same as in Example 4.

Then we have the cumulative hazard rate of a Weibull distribution as

$$\Lambda_0(t) = \int_0^t \theta\mu\zeta^{\mu-1}d\zeta = \theta t^\mu,$$

and

$$\begin{aligned} 1 - H(t|\xi) &= 1 - F(t|\xi) \\ &= \exp(-\theta_0 t^\mu \exp(\beta_0\xi)) \quad \text{for} \quad t \in [0,\tau) \end{aligned} \tag{4.24}$$

and $1 - H(t|\xi) = 0$ for $t \geq \tau$. We define X again to be the two-points covariate with

$$\mathsf{P}(X = \xi_j) = q_j \quad \text{for} \quad j = 1,2, \quad \xi_1, \xi_2 \in [0,1]$$

where $q_1 + q_2 = 1$ and assume that the measure Q is

$$Q(\eta) = \begin{cases} q_j, & \text{if } \eta = \xi_j \quad j = 1,2, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, we obtain with (4.24) and Lemma 4.9

$$\begin{aligned} \mathbf{\Sigma}(\beta_0,\theta_0,\mu,G,Q) &= q_1q_2(\xi_1-\xi_2)^2\exp(\beta_0\xi_1+\beta_0\xi_2)\theta_0\mu\times \\ &\times \int_0^\tau \frac{\exp(-\theta_0u^\mu\exp(\beta_0\xi_1)-\theta_0u^\mu\exp(\beta_0\xi_2))u^{\mu-1}}{q_1\exp(-\theta_0u^\mu\exp(\beta_0\xi_1)+\beta_0\xi_1)+q_2\exp(-\theta_0u^\mu\exp(\beta_0\xi_2)+\beta_0\xi_2)}du \\ &= q_1q_2(\xi_1-\xi_2)^2\exp(\beta_0(\xi_1+\xi_2))\theta_0\mu\times \\ &\times \int_0^\tau \frac{u^{\mu-1}}{q_1\exp(\theta_0u^\mu\exp(\beta_0\xi_2)+\beta_0\xi_1)+q_2\exp(\theta_0u^\mu\exp(\beta_0\xi_1)+\beta_0\xi_2)}du. \end{aligned}$$

The value of $\mathbf{\Sigma}(\beta_0,\theta,\mu,G,Q)$ is computed for several β_0,θ,μ and the values $\mathbf{\Sigma}^*(\beta_0,\theta,\mu,G)$ are given in the Table 4.2.

In the Table 4.2, we find that the local optimal two-points covariates depend on the parameters β_0,θ_0 and μ. The Figure A.4 $a), b)$ and $c)$ illustrate the dependence between $\mathbf{\Sigma}(\beta_0,\theta_0,\mu,G,Q)$ and the ξ_1,ξ_2 for $\beta_0 = 1.0, 3.0, 6.0$, for fixed $q_1 = 0.5$, $\theta_0 = 1$ and $\mu = 2$. The perfectly symmetric plots are shown in Figure A.4 $a), b)$. For both plots, we also see that curves for $\xi_1 = 0$ are strongly increased. In contrast, there are strongly decreasing curves for $\xi_1 = 1$. And we also find that $\mathbf{\Sigma}(1,1,2,G,Q)$ is maximal for $\xi_1^* = 1, \xi_2^* = 0$ or $\xi_1^* = 0, \xi_2^* = 1$.

Table 4.2: Local optimal two-points covariates of the Weibull model (4.23) as a submodel of (3.2).

		$\mu = 2$				$\mu = 5$			
θ	β_0	ξ_1^*	ξ_2^*	q_1^*	$\mathbf{\Sigma}^*(\beta_0, \theta, \mu, G)$	ξ_1^*	ξ_2^*	q_1^*	$\mathbf{\Sigma}^*(\beta_0, \theta, \mu, G)$
1	1	0	1	0.52430	0.20527	0	1	0.52410	0.20521
	2	0	1	0.58355	0.13400	0	1	0.58280	0.13321
	3	0	1	0.64685	0.07686	0	1	0.61970	0.07774
	4	0.1	0.9	0.66075	0.04357	0.2	1	0.64905	0.04640
	5	0.3	0.9	0.64770	0.02801	0.1	0.8	0.67345	0.02946
	6	0.4	1	0.64145	0.02080	0.4	0.9	0.64570	0.02135
	7	0.3	0.8	0.70555	0.01450	0	1	0.64900	0.01614
	8	0.5	1	0.62655	0.01398	0	1	0.67025	0.01314
3	1	0	1	0.52430	0.20527	0	1	0.52550	0.20464
	2	0	1	0.58350	0.13400	0	1	0.57990	0.13204
	3	0	1	0.65170	0.07629	0	1	0.61810	0.08168
	4	0.1	0.9	0.64425	0.04425	0	0.8	0.6051	0.04497
	5	0.3	1	0.61235	0.02951	0.2	0.9	0.64930	0.03162
	6	0.2	0.8	0.66370	0.02081	0.5	1	0.66945	0.02248
	7	0.5	1	0.54085	0.01763	0	1	0.65655	0.01777
	8	0.4	0.8	0.67440	0.01388	0	1	0.60365	0.01357

On the other hand, we do not find the symmetric plots in Figure A.4 c) as $\beta_0 = 6$. The largest value of $\mathbf{\Sigma}(6, 1, 2, G, Q)$ is reached for $\xi_1^* = 1, \xi_2^* = 0.5$ or $\xi_1^* = 0.5, \xi_2^* = 1$.
In Figure A.5, A.6 and A.7 the similar plots are represented with other unknown parameters and the dependence of $\mathbf{\Sigma}(\beta_0, \theta_0, \mu, G, Q)$ and the ξ_1, ξ_2 is obvious. Qualitatively, we have the same behavior as in the exponential models. □

It should be remarked that the values $\mathbf{\Sigma}^*$ do not depend very strongly from λ_0, i.e. here by θ, μ. This property is connected with the special choice of the censoring distribution G.

4.2.2 Sensitivity analysis

The optimality criterion $\det \mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ depends on the unknown parameters $\boldsymbol{\beta}_0 \in \mathbb{R}^p$, $\lambda_0 \in \mathcal{C}([0,\tau])$. The sensitivity analysis will express the relative distinction between the values $\det \mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$ and $\det \mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0))$ for $(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0)$ in a neighborhood of $(\boldsymbol{\beta}_0, \lambda_0)$. In this way the loss of optimality will be measured, if one works with a covariate $Q^*(\boldsymbol{\beta}_0, \lambda_0)$ instead of the optimal $Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0)$. The value

$$\mathrm{eff}_{\tilde{\boldsymbol{\beta}}_0,\tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0)) = \frac{\det \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))}{\det \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0))} \tag{4.25}$$

is called the efficiency of the local optimal covariate $Q^*(\boldsymbol{\beta}_0, \lambda_0)$ under the parameters $(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0)$ in the Cox model, i.e. under the distribution with the hazard rate

$$\lambda(t|\mathbf{x}) = \tilde{\lambda}_0(t)\exp(\tilde{\boldsymbol{\beta}}_0^T \mathbf{x}).$$

This efficiency has to be calculated. Here the special case $p = 1$ and the exponential model is considered. The neighborhood of the exponential model is described by baseline hazard rates in the form

$$\tilde{\lambda}_0(t) = a + b\cos(\omega t)$$

for constants a, b, ω. In the Tables A.1-A.6 the local optimal covariates $Q^*(\tilde{\beta}_0, \tilde{\lambda}_0)$ are given by the support points ξ_i^* and the corresponding weights q_i^*. Moreover, the values

$$\boldsymbol{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G) = \boldsymbol{\Sigma}(\tilde{\beta}_0, \tilde{\lambda}_0, G, Q^*(\tilde{\beta}_0, \tilde{\lambda}_0))$$

and the efficiencies $\mathrm{eff}_{\tilde{\beta}_0,\tilde{\lambda}_0}(Q^*(1,1))$ are calculated. In Tables A.1-A.4 we have chosen $\tilde{\beta}_0 = 1$, in Table A.5 we put $\tilde{\beta}_0 = 2$ and $\tilde{\beta}_0 = 3$ is used in Table A.6-A.9. In the Tables the values of efficiencies are given for $(\boldsymbol{\beta}_0, \lambda_0) = (1, 1)$ and $\tilde{\lambda}_0$ with $a = 1, 2, 5$ and $b = 0, 0.3, 0.6, 0.9$ and $\omega = 1$. Moreover, the efficiencies are calculated for $\tilde{\lambda}_0$ with $\omega = 15$, $b = 0.75$ and $\omega = 30$, $b = 0.9$.

Remarkable is the fact that for fixed $\tilde{\boldsymbol{\beta}}_0$ the efficiencies do not vary very much for different values of $\tilde{\lambda}_0$.

The efficiencies are defined for a given censoring distribution. We are interested in the influence of G for these efficiencies. The important consideration is generating censoring time with some specific distribution. Now, we choose four different censoring distributions. The first censoring distribution is a censoring where one censors at a fixed time τ. The second distribution describes a G where one censors in an exponential way and here mainly at the beginning. The third distribution demonstrate that censored observations occur most often at nearly time τ and the last censoring distribution describes a constant censoring.

More precisely, we compare the values of $\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0))$ for the following censoring distributions,

1. $G(t) = 1 - \delta_{[0,\tau]}(t)$

2. $G(t) = 1 - \exp(-t)$

3. $G(t) = (\exp(ct) - 1)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}\log 2$

4. $G(t) = (ct)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}$.

From Tables A.1, A.5 and A.6, we note that for each fixed $\tilde{\lambda}_0$ and censoring distribution $G(t) = 1 - \delta_{[0,\tau)}(t)$, the asymptotic variances $\mathbf{\Sigma}^{-1}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0))$ will increase with increasing $\tilde{\boldsymbol{\beta}}_0$.

For each fixed $\tilde{\lambda}_0$ and $\tilde{\boldsymbol{\beta}}_0$, the asymptotic variances $\mathbf{\Sigma}^{-1}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0))$ dictates that there is a big difference between the results under second and third censoring distribution. The second censoring distribution $G(t) = 1 - \exp(-t)$ will lead to much censoring observations in the data set at the beginning of the study, which will increase the loss of information in the data. On the other hand, the third censoring pattern $G(t) = (\exp(ct) - 1)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \dfrac{1}{\tau}\log 2$ indicates that there is not much censoring in the beginning of the study, so that we have more enough uncensored lifetimes to assess in this case.
Thus, from Table A.2 and A.3 we can see clearly that the values of $\mathbf{\Sigma}^{-1}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0))$ under second censoring distribution is higher than under third censoring distribution.
Moreover, for cases with the fourth censoring distribution $G(t) = (ct)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \dfrac{1}{\tau}$, the difference of the values of $\mathbf{\Sigma}^{-1}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0))$ corresponding to fixed $\tilde{\boldsymbol{\beta}}_0$ and $\tilde{\lambda}_0$ are almost the same as under the third censoring distribution. These results are shown in Table A.4.

4.3 Proposals for applicable optimal covariates

In the previous section, local D-optimal covariates were defined and they were calculated in examples. Some procedures exist for getting optimal covariates which do not depend explicitly on unknown parameters. In this section, we consider two-stage optimal covariates, which are also called estimated optimal covariates. Here the sample is divided into two parts. With the observations from the first part the unknown parameters are estimated, these estimations are substituted for the unknown parameter functional $\det \mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$, getting $\det \mathbf{\Sigma}(\widehat{\boldsymbol{\beta}}_0, \widehat{\lambda}_0, G, Q)$ and this estimated function will be maximized with respect to Q.
A second kind of applicable D-optimal covariates bases on the continuity of $\det \mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$. We calculate optimal covariates (measure Q^*) for a fixed $(\boldsymbol{\beta}_0, \lambda_0(\cdot))$ and then we know that Q^* is nearly optimal for all $(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0)$, which are in a small neighborhood around $(\boldsymbol{\beta}_0, \lambda_0)$. A third kind of applicable D-optimal covariates is connected with a grid $\{(\boldsymbol{\beta}_{0i}, \lambda_{0i}), i = 1, \ldots, r\}$ in the parameter space and a mixture of the $\det \mathbf{\Sigma}(\boldsymbol{\beta}_{0i}, \lambda_{0i}, G, Q)$ is to be maximized.

4.3.1 Two-stage choice of optimal covariates

We defined D-optimal covariates and we have realized that these local optimal covariates depend on the $\boldsymbol{\beta}_0$ and on the baseline hazard rate and on G. For getting realistic nearly optimal covariates we propose a two-stage design to solve the estimation problem. We divide the observations in two parts. The first part of observations, which is $n_1 = \nu n, 0 < \nu < 1$, are taken from the model with covariates described by a measure Q is not optimal, so that Q will be chosen as a reasonable measure, possibly practical experiences are to be used. With these observations we estimate $\boldsymbol{\beta}_0$, λ_0 and Λ_0. The estimator $\widehat{\boldsymbol{\beta}}_0$ will be the maximum partial likelihood estimate, $\widehat{\Lambda}_0$ will be the Breslow estimate.
Then the second part of observations, namely $n_2 = (1-\nu)n$ observations are observed according to a local optimal covariate, namely locally to $(\widehat{\boldsymbol{\beta}}_0, \widehat{\lambda}_0)$.

Definition 4.5. *We call a measure $\widehat{Q}$ estimated D-optimal in $\mathcal{D}$ if*

$$\det \mathbf{\Sigma}(\widehat{\boldsymbol{\beta}}_0, \widehat{\lambda}_0, G, \widehat{Q}) = \max_{Q \in \mathcal{D}} \det \mathbf{\Sigma}(\widehat{\boldsymbol{\beta}}_0, \widehat{\lambda}_0, G, Q).$$

The covariate $\widehat{\boldsymbol{X}}$ with the induced measure $\widehat{Q}$ is called estimated D-optimal covariate.

The resulting measure

$$\tilde{Q} = \nu Q + (1 - \nu)\widehat{Q}$$

is called a two-stage D-optimal measure for choosing covariates or we say that the variable $\tilde{\boldsymbol{X}}$ with the induced measure $\tilde{Q}$ is two-stage D-optimal. By changing the rate ν, we can change the relations of the sample sizes in the two stages.

4.3.2 Nearly D-optimal covariates

Lemma 4.8 states the continuity of the matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ is continuous in all arguments and if for fixed $(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ the maximizer Q^* with

$$Q^* = \arg\max_{Q} \det \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q) = Q^*(\boldsymbol{\beta}_0, \lambda_0)$$

is unique, then the continuity of $\boldsymbol{\Sigma}$ implies that for any $\epsilon > 0$ there exists a $\delta > 0$ with

$$\|Q^*(\boldsymbol{\beta}_0, \lambda_0) - Q^*(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0)\| < \epsilon,$$

$$|\max_{Q} \det \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q) - \max_{Q} \det \boldsymbol{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q)| < \epsilon,$$

if

$$|\boldsymbol{\beta}_0 - \tilde{\boldsymbol{\beta}}_0| < \delta, \quad \|\lambda_0(\cdot) - \tilde{\lambda}_0(\cdot)\| < \delta.$$

This means for $\tilde{\boldsymbol{\beta}}_0$ in a neighborhood of $\boldsymbol{\beta}_0$ and $\tilde{\lambda}_0(\cdot)$ in a tube around $\lambda_0(\cdot)$ the calculated $Q^*(\boldsymbol{\beta}_0, \lambda_0)$ is nearly D-optimal. The tube around λ_0 is a nonparametric set, the δ-neighborhood around $\boldsymbol{\beta}_0$ is a parametric set. The MPLE $\widehat{\boldsymbol{\beta}}$ which is a estimator of $\boldsymbol{\beta}_0$ can be used for all $(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0)$ in the given neighborhood. The covariates $\boldsymbol{X}^*$ with the induced measure $Q^*(\boldsymbol{\beta}_0, \lambda_0)$ are called nearly D-optimal covariates.

4.3.3 Mixed optimal covariates

To propose a further methods for choosing covariates in a nearly optimal way, let us introduce a grid in the set $\mathcal{F} \subseteq \mathbb{R}^p \times \mathcal{C}([0,\tau])$. $\mathcal{F}$ denotes a set of models, describes by $(\boldsymbol{\beta}_0, \lambda_0) \in \mathbb{R}^p \times \mathcal{C}([0,\tau])$. The grid is denoted by

$$(\boldsymbol{\beta}_{0i}, \lambda_{0i}(\cdot)), \quad i = 1, \ldots, r$$

and we define as optimality criterion the mixed functional of the form

$$M(Q) = \sum_{i=1}^{r} p_i \det \boldsymbol{\Sigma}(\boldsymbol{\beta}_{0i}, \lambda_{0i}, G, Q) \tag{4.26}$$

for positive constants $p_1, \ldots, p_r$ which express the importance of the grid points

$$(\boldsymbol{\beta}_{01}, \lambda_{01}), \ldots, (\boldsymbol{\beta}_{0r}, \lambda_{0r}).$$

The maximizer of the functional M, say Q^*, is called “mixed optimal covariates”. This Q^* is not local optimal for any point in $\mathcal{F}$ but quite well for the most of the points.

4.4 Suboptimal covariates

In general, one is looking for optimal covariates but the determination or calculation is impossible in many problems. Therefore, a weaker optimality is discussed in classical experimental design problems . One determines such designs, which are not improvable by changes in a given class. This restricted optimality is called ”suboptimality” and will be discussed here in the context of the covariates.

In the former subsections several optimality criteria were introduced. For the local optimal covariates the criterion

$$\det \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q) \tag{4.27}$$

was the basis for finding an optimal design. The two-stage optimality criterion was principally the same, here the $\boldsymbol{\beta}_0, \lambda_0$ were substituted by estimates. For the nearly optimal covariates this functional (4.27) was used and for the mixed strategy we worked with (4.26). Now all these criteria are denoted by $\tilde{M} = \tilde{M}(Q)$. Let $Q_{\boldsymbol{x}}$ be the one-point measure with

$$Q_{\boldsymbol{x}}(\eta) = \begin{cases} 1, & \text{if } \boldsymbol{\eta} = \boldsymbol{x}, \\ 0, & \text{otherwise.} \end{cases}$$

We are looking for D-optimal covariates $\boldsymbol{X}^*$, i.e. we look for a measure Q^* with

$$\det \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q^*) = \max_Q \det \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q) \tag{4.28}$$

and Q is any measure over $\boldsymbol{\mathcal{X}} = [0,1]^p$. For the described variations of criteria we are looking for a Q^* with

$$\tilde{M}(Q^*) = \max_Q \tilde{M}(Q).$$

This is the same as

$$\tilde{M}(Q^*) \geq \tilde{M}((1-\alpha)Q^* + \alpha Q) \tag{4.29}$$

for all measures Q and for all $\alpha \in [0,1]$. Whittle (1973) proved for concave functionals $\tilde{M}$, i.e. for $\tilde{M}$ with

$$\tilde{M}((1-\alpha)Q_1 + \alpha Q_2) \geq (1-\alpha)\tilde{M}(Q_1) + \alpha \tilde{M}(Q_2)$$

for any measures Q_1, Q_2 and for all $\alpha \in [0,1]$, the condition (4.29) is equivalent to

$$\tilde{M}(Q^*) \geq \tilde{M}((1-\alpha)Q^* + \alpha Q_{\boldsymbol{x}}) \tag{4.30}$$

for any one-point measure $Q_{\boldsymbol{x}}$, $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$, $\alpha \in [0,1]$. Moreover, for concave functionals $\tilde{M}$ this relation (4.30) is equivalent to

$$\frac{d}{d\alpha}\tilde{M}((1-\alpha)Q^* + \alpha Q_{\boldsymbol{x}})|_{\alpha=0} \leq 0 \tag{4.31}$$

for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. Hence, one has for concave functionals $\tilde{M}$, the property that Q^* from (4.29) is determined by the condition (4.31). The importance of condition (4.31) consists in a good possibility for numerical checking. In this way we will define suboptimal covariates.

Definition 4.6. *Suppose that $\boldsymbol{\beta}_0$ is asymptotically estimable. The measure Q^* is called D-suboptimal if*

$$\frac{d}{d\alpha}\tilde{M}((1-\alpha)Q^* + \alpha Q_{\boldsymbol{x}})|_{\alpha=0} \leq 0 \tag{4.32}$$

for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. $\tilde{M}$ stands here for any criterion mentioned above. The corresponding covariate $\boldsymbol{X}^*$ with the induced measure Q^* is called D-suboptimal covariate.

Hence, any D-optimal covariate is a D-suboptimal covariate. In classical experimental design problems the suboptimal designs are optimal designs, at least for concave criteria. Suboptimal covariates have the advantage that there exist iterative procedures to calculate these covariates. Moreover, suboptimality can be checked.

4.4.1 Iterative calculation of suboptimal covariates

The inequality (4.32) gives the idea for an iteration process. In experimental design iteration processes for calculating D-optimal designs were proposed by Kiefer (1961), Wynn (1972); therefore, we can use a similar procedure here. The essential difference to the classical design problems consists in the property of the asymptotic variance matrix. We cannot prove that $\det \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ is concave with respect to Q.
We denote with

$$\tilde{Q} = (1-\alpha)Q_{(0)} + \alpha Q_{\boldsymbol{x}}, \quad \alpha \in [0,1] \tag{4.33}$$

where a starting measure $Q_{(0)}$ is chosen in such a way that $\boldsymbol{\beta}_0$ is asymptotically estimable.
From the function

$$\frac{\partial}{\partial\alpha}\tilde{M}(\tilde{Q})|_{\alpha=0} =: \phi(Q_{(0)}, \boldsymbol{x}),$$

the dependence on $\boldsymbol{\beta}_0$, λ_0 will be suppressed in ϕ. $\boldsymbol{\mathcal{X}} \in \mathbb{R}^p$ denotes the set of possible values of the covariates. Let us formulate the iteration procedure.

Step 1 Choose a measure $Q_{(0)}$ in such a way that $\boldsymbol{\beta}_0$ is asymptotically estimable.

Step 2 Choose such a point $\boldsymbol{x}_0 \in \boldsymbol{\mathcal{X}}$ that

$$\phi(Q_{(0)}, \boldsymbol{x}_0) = \max_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} \phi(Q_{(0)}, \boldsymbol{x}).$$

Step 3 Choose an α_0 with $0 < \alpha_0 < 1$ in such a way that

$$\tilde{M}((1-\alpha_0)Q_{(0)} + \alpha_0 Q_{\boldsymbol{x}_0}) = \max_{\alpha} \tilde{M}((1-\alpha)Q_{(0)} + \alpha Q_{\boldsymbol{x}_0}).$$

We put $Q_{(1)} = (1-\alpha_0)Q_{(0)} + \alpha_0 Q_{\boldsymbol{x}_0}$.
These steps will be repeated, now starting with $Q_{(1)}$, finding $\boldsymbol{x}_1$ with

$$\phi(Q_{(1)}, \boldsymbol{x}_1) = \max_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} \phi(Q_{(1)}, \boldsymbol{x}).$$

As in step 3 we determine an $\alpha_1 \in (0,1)$ in such a way that

$$\tilde{M}((1-\alpha)Q_{(1)} + \alpha Q_{\boldsymbol{x}_1})$$

is maximal. In this way we get a sequence of measures $Q_{(0)}, Q_{(1)}, Q_{(2)}, \ldots$ with

$$\tilde{M}(Q_{(0)}) \leq \tilde{M}(Q_{(1)}) \leq \tilde{M}(Q_{(2)}) \leq \ldots.$$

Then the measures $Q_{(i)}$ are improved stepwise. The criterion $\tilde{M}$ and the functional ϕ can be calculated very easily. The iteration stops if $\tilde{M}(Q_{(l)})$ is not improvable in this way or if the improvement is smaller than a given bound.
In the criteria $\tilde{M}$ the term $\det \boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ is contained, either as a single part or in a weighted sum. We recall the representation of Σ as in Theorem 4.7. For measures Q_0, Q_1, Q and $\mu \in \mathbb{R}$

$$\begin{aligned} A_1(Q,\mu) &:= \int \exp\Big(-\mu^- e^{\boldsymbol{\beta}_0^T \boldsymbol{\eta}} + \boldsymbol{\beta}_0^T \boldsymbol{\eta}\Big) dQ(\boldsymbol{\eta}) \\ A_2(Q_0,Q_1,\mu) &:= \int\int \exp\Big(-\mu^- e^{\boldsymbol{\beta}_0^T \boldsymbol{\zeta}} + \boldsymbol{\beta}_0^T \boldsymbol{\zeta}\Big) \exp\Big(-\mu^- e^{\boldsymbol{\beta}_0^T \boldsymbol{\rho}} + \boldsymbol{\beta}_0^T \boldsymbol{\rho}\Big) \times \\ &\quad \times (\boldsymbol{\zeta} - \boldsymbol{\rho})(\boldsymbol{\zeta} - \boldsymbol{\rho})^T \Big(1 - G(\Lambda_0^{-1}(\mu^-)\Big) dQ_0(\boldsymbol{\zeta}) dQ_1(\boldsymbol{\rho}). \end{aligned}$$

Then we have for $\tilde{Q} = (1-\alpha)Q_{(0)} + \alpha Q$

$$\begin{aligned} &\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, \tilde{Q}) \\ &= \int_0^{\Lambda_0(\tau)} \frac{(1-\alpha)^2 A_2(Q_{(0)}, Q_{(0)}, u) + 2\alpha(1-\alpha) A_2(Q_{(0)}, Q, u) + \alpha^2 A_2(Q, Q, u)}{(1-\alpha) A_1(Q_{(0)}, u) + \alpha A_1(Q, u)} \, du. \end{aligned}$$

Therefore, we get

$$\frac{\partial}{\partial\alpha}\mathbf{\Sigma}(\boldsymbol{\beta}_0,\lambda_0,G,\tilde{Q})|_{\alpha=0}$$
$$=\int_0^{\Lambda_0(\tau)}\frac{1}{[A_1(Q_{(0)},u)]^2}\Bigg\{\Big\{A_1(Q_{(0)},u)\{-2A_2(Q_{(0)},Q_{(0)},u)+2A_2(Q_{(0)},Q,u)\}\Big\}$$
$$-\Big\{A_2(Q_{(0)},Q_{(0)},u)\{-A_1(Q_{(0)},u)+A_1(Q,u)\}\Big\}\Bigg\}du$$
$$=-\mathbf{\Sigma}(\boldsymbol{\beta}_0,\lambda_0,G,Q_{(0)})+2\int_0^{\Lambda_0(\tau)}\frac{A_2(Q_{(0)},Q,u)}{A_1(Q_{(0)},u)}du-\int_0^{\Lambda_0(\tau)}\frac{A_1(Q,u)A_2(Q_{(0)},Q_{(0)},u)}{[A_1(Q_{(0)},u)]^2}du.$$

In the Example 6 these representations are calculated for an m-point measure $Q_{(0)}$ and one-point measure $Q = Q_{\boldsymbol{x}}$

Example 6. We consider the Cox model with $p = 1$. Let $Q_{(0)}$ be defined as an m-point measure as in Theorem 4.4. Let the censoring G be as in Example 4. Then we have

$$\mathbf{\Sigma}(\boldsymbol{\beta}_0,\lambda_0,G,Q_{(0)})=\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{m}q_iq_je^{\beta_0(\xi_i+\xi_j)}(\xi_i-\xi_j)^2\int_0^{\Lambda_0(\tau)}\frac{e^{-u(e^{\beta_0\xi_i}+e^{\beta_0\xi_j})}}{\sum_{l=1}^{m}q_le^{\beta_0\xi_l}e^{(-ue^{\beta_0\xi_l})}}du. \tag{4.34}$$

With an additional one-point measure Q_x the convex linear combination

$$\tilde{Q}=(1-\alpha)Q_{(0)}+\alpha Q_x.$$

will be considered. With

$$g_i(u,\beta_0,\lambda_0)=q_i\exp(-ue^{\beta_0\xi_i}+\beta_0\xi_i)$$

and with (4.10) we have

$$\begin{aligned}\mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, \tilde{Q}) =& \frac{1}{2}(1-\alpha)^2 \sum_{j=1}^{m}\sum_{i=1}^{m} q_i q_j e^{\beta_0(\xi_i+\xi_j)}(\xi_i-\xi_j)^2 \times \\ & \times \int_0^{\Lambda_0(\tau)} \frac{\exp(-u(e^{\beta_0\xi_i}+e^{\beta_0\xi_j}))}{(1-\alpha)\sum_{l=1}^{m} g_l(u,\beta_0,\lambda_0)+\alpha\exp(-ue^{\beta_0 x}+\beta_0 x)} du \\ & + \alpha(1-\alpha)\sum_{j=1}^{m} q_j e^{\beta_0(\xi_j+x)}(\xi_j-x)^2 \times \\ & \times \int_0^{\Lambda_0(\tau)} \frac{\exp(-u(e^{\beta_0\xi_j}+e^{\beta_0 x}))}{(1-\alpha)\sum_{l=1}^{m} g_l(u,\beta_0,\lambda_0)+\alpha\exp(-ue^{\beta_0 x}+\beta_0 x)})du.\end{aligned}$$

Now we put $p=1$ and because of $\Sigma = det\Sigma$ the function $\phi(Q_{(0)}, x)$ has the following form:

$$\begin{aligned}\phi(Q_{(0)}, x) =& \frac{\partial}{\partial\alpha}\mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, \tilde{Q})|_{\alpha=0} \\ =& -\mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q_{(0)}) \\ & + \int_0^{\Lambda_0(\tau)} \frac{\exp(-ue^{\beta_0 x}+\beta_0 x)\sum_{i=1}^{m} g_i(u,\beta_0,\lambda_0)(\xi_i-x)^2}{\sum_{l=1}^{m} g_l(u,\beta_0,\lambda_0)} du \\ & - \frac{1}{2}\int_0^{\Lambda_0(\tau)} \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} g_i(u,\beta_0,\lambda_0) g_j(u,\beta_0,\lambda_0)(\xi_i-\xi_j)^2}{\left(\sum_{l=1}^{m} g_l(u,\beta_0,\lambda_0)\right)^2} du.\end{aligned}$$

These explicit representations of $\mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, \tilde{Q})$ and $\phi(Q_{(0)}, x)$ are used for the iterative calculation of suboptimal covariates. For calculating local D-suboptimal covariates for exponential models as submodels of (3.2), we start the iterative procedures with the optimal two-points covariates given in Table 4.1 or some other appropriate designs. We demonstrate this for $\lambda_0 = 1$ and choose the parameters $\beta_0 = 1, 1.5, 2, \ldots, 6$ step by step. For $\beta_0 = 1$, we start with $Q_{(0)}$ and

$$Q_{(0)}(0) = 0.52431, \quad Q_{(0)}(1) = 0.47569.$$

Then we calculate $\phi(Q_{(0)}, x)$ and find that $\phi(Q_{(0)}, x) \leq 0$ for $x \in [0,1]$. Consequently $Q_{(0)}$ is suboptimal if $\beta_0 = 1$.
For $\beta_0 = 6$ we start with $Q_{(0)}$ and

$$Q_{(0)}(0.4) = 0.67585, \quad Q_{(0)}(1) = 0.32415.$$

Then we calculate $\phi(Q_{(0)}, x)$ and find that $\max_{x\in[0,1]} \phi(Q_{(0)}, x)$ is reached in $x = 0$, the corresponding α_0 from step 3 in the iteration procedure is $\alpha_0 = 0.359$. We have $Q_{(1)}$ with

$$Q_{(1)}(0.4) = 0.4332, \quad Q_{(1)}(0) = 0.359, \quad Q_{(1)}(1) = 0.2078.$$

After getting three values of $Q_{(1)}$, $\phi(Q_{(1)}, x)$ is calculated and the maximum of $\phi(Q_{(1)}, x)$ is reached for $x = 0.621$. The value α_1 equals to 0.09033. Thus, we obtain the measure $Q_{(2)}$ with

$$Q_{(2)}(0) = 0.32657, \quad Q_{(2)}(0.400) = 0.39407,$$

$$Q_{(2)}(0.621) = 0.09033, \quad Q_{(2)}(1) = 0.18903.$$

This measure $Q_{(2)}$ is considered as the suboptimal covariates.
We notice that some rounding errors occurred in the numerical calculations and for the most of the values x, the values $\phi(Q_{(2)}, x)$ are negative. Only for some $\tilde{x}$ it can be happened that $\phi(Q_{(2)}, \tilde{x}) \approx 0$, but positive. Then one checks that $\mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, (1-\alpha)Q_{(2)} + \alpha Q_{\tilde{x}}) < \mathbf{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$ and it is justified to consider $Q_{(2)}$ as suboptimal. Here we described the iteration for $\beta_0 = 1$, $\beta_0 = 6$ as example.

Table 4.3: Local suboptimal covariates for the exponential model as a submodel of (3.2)

β_0	ξ_1^*	ξ_2^*	ξ_3^*	ξ_4^*	q_1^*	q_2^*	q_3^*	q_4^*	$\mathbf{\Sigma}^*(\beta_0, \lambda_0, G, Q)$
1	0	1			0.52431	0.47569			0.20525
1.5	0	1			0.55200	0.44800			0.16915
2	0	1			0.58355	0.41645			0.13400
3	0	0.525	1		0.54485	0.15867	0.29648		0.08075
4	0.1	0.535	0.900	1	0.48093	0.17499	0.25075	0.09333	0.04845
5	0	0.200	0.565	0.900	0.28000	0.36616	0.16584	0.18800	0.03665
6	0	0.400	0.621	1	0.32657	0.39407	0.09033	0.18903	0.03205

In Table 4.3 local suboptimal covariates are given. It can be clearly seen that the suboptimal covariates for $\boldsymbol{\beta}_0 = 1, 1.5, 2$ coincide with the local optimal two-points covariates in Table 4.1. For larger $\boldsymbol{\beta}_0$, we see by comparing the Tables 4.1 and 4.3 that three- or four-point covariates lead to higher values of $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \lambda_0, G, Q)$, i.e. if $Q_{\beta_0,1}$ describes the measures from Table 4.1, and $Q_{\beta_0,2}$ describes the measures from Table 4.3, then we have

$$\boldsymbol{\Sigma}(i, 1, G, Q_{i,1}) < \boldsymbol{\Sigma}(i, 1, G, Q_{i,2}) \quad i = 3, 4, 5, 6.$$

This means that for $\beta_0 = 1, 1.5, 2$ two-points covariates are local suboptimal. For $\beta_0 = 3, 4, 5, 6$ the local optimal two-points covariates are not local suboptimal in the set of all covariates because we found three- or four-point covariates with a smaller asymptotic variance.

Chapter 5

Statistical inference in the Cox model with time-dependent coefficient

In this chapter we consider the extension of the Cox model where the coefficients are allowed to depend on time. Such models are useful to describe situation where the effect of the covariate–not the covariate itself–varies with the time.

As already mentioned in (3.17), the model is given by

$$\lambda(t, \boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T(t)\boldsymbol{x}(t)),$$

where $\boldsymbol{\beta}(\cdot)$ is a p-dimensional vector of time-dependent coefficients function.

Here we will concentrate on estimating and testing the coefficient function $\boldsymbol{\beta}$, so that for simplicity of presentation we suppose that the covariate does not depend on time, i.e. $\boldsymbol{x}(t) = \boldsymbol{x}$.

In the first section of this chapter the local partial partial likelihood estimation method is described, here we follow namely the approach of Cai and Sun (2003). The main topic of this chapter is a test for checking the parametric form of the coefficient function $\boldsymbol{\beta}$. The test procedure of this chapter, which is presented in Section 5.2 is based on the score function.

5.1 The maximum local partial likelihood estimation

The idea of the local partial likelihood estimation in the Cox model goes back to the general method of smoothing by local polynomials. Using a Taylor expansion the function of interest is approximated locally. The simplest special case is a local approximation by a constant, the mostly applied method is the local linear approximation. The local approximation is coupled with the log partial likelihood function, the local neighborhood is described by a kernel function and a bandwidth sequence.
The local partial likelihood method for the estimation in hazard regression was considered by Hastie and Tibshirani (1993). They considered the estimation of the function $\psi(x)$ in the model $\lambda(t|x) = \lambda_0(t)\exp(\psi(x))$ and proposed a nearest neighbor type estimator with uniform windows.
Fan et al. (1997) investigated a more general approach and compared local likelihood and local partial likelihood estimators. Cai and Sun (2003) used this local linear partial likelihood approach to estimate the time-dependent coefficient function $\boldsymbol{\beta}$ in model (3.17) and also established the asymptotic consistency and normality of the estimators $\widehat{\boldsymbol{\beta}}_n(t)$ at a fixed point t.
In this section we will explain this method and illustrate this by an example. Further, since for the testing procedure proposed in the next section, we need the method based on the local constant method, we will also present this method.
We suppose the extended Cox model with the hazard rate

$$\lambda(t, \boldsymbol{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T(t)\boldsymbol{x}). \tag{5.1}$$

The logarithm of the partial likelihood function depending on the function $\boldsymbol{\beta}(\cdot)$ is given by

$$\ell_n(\boldsymbol{\beta}) = \sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{\beta}^T(u)\boldsymbol{X}_i - \log\left\{\sum_{j=1}^{n} Y_j(u)\exp(\boldsymbol{\beta}^T(u)\boldsymbol{X}_j)\right\}\right]dN_i(u). \tag{5.2}$$

Of course, this can not be the basis to determine an estimator of the function $\boldsymbol{\beta}$. The idea for the maximum local linear partial likelihood estimation is to replace the function $\boldsymbol{\beta}$ in (5.2) locally by the first two terms of its Taylor expansion.

For s in a neighborhood of t we have for the component β_j

$$\beta_j(s) \approx \beta_j(t) + \beta_j'(t)(s-t).$$

where β_j' is the first derivative of β_j.
Set $b_{1j}(t) = \beta_j(t)$ and $b_{2j}(t) = \beta_j'(t)$, then a linear approximation at s is given by

$$\beta_j(s) \approx b_{1j}(t) + b_{2j}(t)(s-t).$$

The local neighborhood of s is characterized by the kernel function $K : \mathbb{R} \to \mathbb{R}$ and a bandwidth sequence $h = h_n$, which controls the size of the neighborhood.
Let $\tilde{\boldsymbol{X}}_i(s, s-t)$ be a $2p$-dimensional vector with

$$\tilde{\boldsymbol{X}}_i(u, u-t) = \boldsymbol{X}_i \otimes (1, u-t)^T,$$

where $\otimes$ denotes the Kronecker product.
The log local linear partial likelihood function for estimating the true parameter function $\boldsymbol{\beta}_0$ at the grid point t is given by

$$\begin{aligned} \ell_t(\boldsymbol{b}) = & \sum_{i=1}^{n} \int_0^{\tau} K_h(u-t) \Bigg[\boldsymbol{b}^T \tilde{\boldsymbol{X}}_i(u, u-t) \\ & - \log \Big\{ \sum_{j=1}^{n} Y_j(u) \exp(\boldsymbol{b}^T \tilde{\boldsymbol{X}}_j(u, u-t)) \Big\} \Bigg] dN_i(u) \end{aligned} \tag{5.3}$$

where $\boldsymbol{b} = (b_{11}(t), \ldots, b_{1p}(t), b_{21}(t), \ldots, b_{2p}(t))^T$, $K_h(s-t) = \frac{1}{h} K\left(\frac{s-t}{h}\right)$ and K is a kernel function satisfying some regularity conditions specified later.
In other words, for each t one has to compute an estimate similarly as defined in the model with time-dependent coefficients; of course in addition we have to take into account the kernel weights.

Let $\widehat{\boldsymbol{b}} = (\widehat{b}_{11}(t), \ldots, \widehat{b}_{1p}(t), \widehat{b}_{21}(t), \ldots, \widehat{b}_{2p}(t))^T$ be the maximizer of local linear partial likelihood function (5.3). Then, $\widehat{\boldsymbol{\beta}}_n(t) = (\widehat{b}_{11}(t), \ldots, \widehat{b}_{1p}(t))^T$ is a maximum local linear partial likelihood estimator for the coefficients function $\boldsymbol{\beta}_0(\cdot)$ at point t.

For the derivation of the maximum we compute as usual the score function and the Hessian matrix. Let us now describe the way of computing.

We use a similar notation as in Chapter 3

$$S_{0n}(u,\boldsymbol{b},t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\exp(\boldsymbol{b}^T\tilde{\boldsymbol{X}}_i(u,u-t)),$$
$$\mathbf{S}_{1n}(u,\boldsymbol{b},t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\tilde{\boldsymbol{X}}_i(u,u-t)\exp(\boldsymbol{b}^T\tilde{\boldsymbol{X}}_i(u,u-t)),$$
$$\mathbf{S}_{2n}(u,\boldsymbol{b},t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\tilde{\boldsymbol{X}}_i(u,u-t)\tilde{\boldsymbol{X}}_i(u,u-t)^T\exp(\boldsymbol{b}^T\tilde{\boldsymbol{X}}_i(u,u-t)).$$

Then $\ell_t(\boldsymbol{b})$ can be written in the form

$$\ell_t(\boldsymbol{b}) = \sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\left[\boldsymbol{b}^T\tilde{\boldsymbol{X}}_i(u,u-t) - \log\left\{nS_{0n}(u,\boldsymbol{b},t)\right\}\right]dN_i(u)$$

and immediately we obtain the score function which is a $2p$-dimensional vector

$$\mathbf{U}_n(\boldsymbol{b}) = \sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\left[\tilde{\boldsymbol{X}}_i(u,u-t) - \frac{\mathbf{S}_{1n}(u,\boldsymbol{b},t)}{S_{0n}(u,\boldsymbol{b},t)}\right]dN_i(u). \qquad (5.4)$$

The Hessian matrix is given by

$$-\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\left[\frac{\mathbf{S}_{2n}(u,\boldsymbol{b},t)}{S_{0n}(u,\boldsymbol{b},t)} - \left\{\frac{\mathbf{S}_{1n}(u,\boldsymbol{b},t)}{S_{0n}(u,\boldsymbol{b},t)}\right\}^{\otimes 2}\right]dN_i(u).$$

Let us consider this in a more detail because the Hessian matrix yields also the method for the computation of $\widehat{\boldsymbol{\beta}}_n(t)$.

$$S_{i0n}(u,\boldsymbol{b},t) = \exp(\boldsymbol{b}^T\tilde{\boldsymbol{X}}_i(u,u-t)).$$

Thus, the Hessian matrix has the form

$$\ell_t''(\boldsymbol{b}) = -\sum_{i=1}^{n}\int_0^{\tau}\frac{K_h(u-t)}{S_{0n}^2(u,\boldsymbol{b},t)}\left[\sum_{j<k} Y_j(u)Y_k(u)S_{j0n}(u,\boldsymbol{b},t)S_{k0n}(u,\boldsymbol{b},t)\right.$$
$$\left.\times(\boldsymbol{X}_j - \boldsymbol{X}_k)^{\otimes 2}\otimes\begin{pmatrix}1 & u-t\\ u-t & (u-t)^2\end{pmatrix}\right]dN_i(u). \qquad (5.5)$$

We can see that the right-hand side of this equation is negative definite as $n \to \infty$. This implies that log local linear partial likelihood $\ell_t(\boldsymbol{b})$ is a strictly concave function of $\boldsymbol{b}$ and has a unique maximum.
The computation of the value of the estimator can be obtained by using the Newton-Raphson algorithm in practice.
The $(j+1)$th Newton-Raphson iteration equation is

$$\widehat{\boldsymbol{b}}_n^{(j+1)} = \widehat{\boldsymbol{b}}_n^{(j)} - \left\{ l''(\widehat{\boldsymbol{b}}_n^{(j)}) \right\}^{-1} l'(\widehat{\boldsymbol{b}}_n^{(j)}) \tag{5.6}$$

where $\widehat{\boldsymbol{b}}_n^{(j)}$ is jth iteration.

To derive the local constant estimator we include only the first term of the Taylor expansion, i.e. $\beta_j(s) \approx b_{1j}(t)$ in the neighborhood of s consequently the log local constant partial likelihood function to be maximized is

$$\ell_t(\boldsymbol{b}) = \sum_{i=1}^{n} \int_0^{\tau} K_h(u-t) \left[\boldsymbol{b}^T \boldsymbol{X}_i - \log \left\{ \sum_{j=1}^{n} Y_j(u) \exp(\boldsymbol{b}^T \boldsymbol{X}_j) \right\} \right] dN_i(u). \tag{5.7}$$

The functions introduced for the computation of the estimator become simpler

$$\begin{aligned} S_{0n}(u, \boldsymbol{b}) &= \sum_{i=1}^{n} Y_i(u) \exp(\boldsymbol{b}^T \boldsymbol{X}_i), \\ \mathbf{S}_{1n}(u, \boldsymbol{b}) &= \sum_{i=1}^{n} Y_i(u) \boldsymbol{X}_i \exp(\boldsymbol{b}^T \boldsymbol{X}_i), \\ \mathbf{S}_{2n}(u, \boldsymbol{b}) &= \sum_{i=1}^{n} Y_i(u) \boldsymbol{X}_i \boldsymbol{X}_i^T \exp(\boldsymbol{b}^T \boldsymbol{X}_i). \end{aligned}$$

with $\boldsymbol{b} = (b_{11}(t), \ldots, b_{1p}(t))^T$.

5.1.1 Simulation Study

According to Cai and Sun (2003), their numerical studies were carried out using Fortran77 and the exact details of the algorithm were not given. Thus an own procedure for the realization of the maximum local partial likelihood estimates were developed. The R-codes used to carry out the simulation

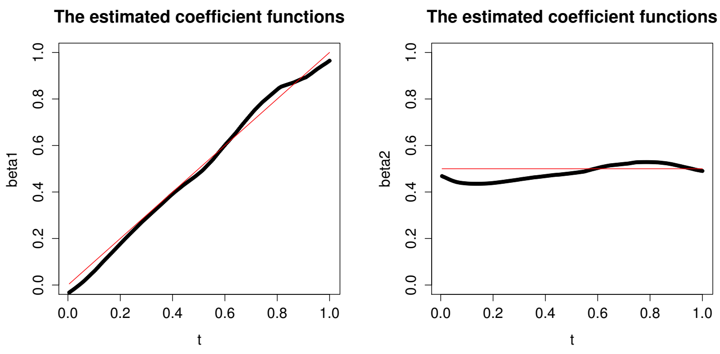
and the basic idea is as follows: We start by generating data and write the log local constant and local linear partial likelihood functions. The estimator based on both functions were obtained using the default setting of the `maxLik` function in the R software.
In this section we illustrate the estimates for two scenarios. The first corresponds to the maximum local linear partial likelihood approach with two covariates. The second considers especially the maximum local constant partial likelihood approach with one covariate. The Epanechnikov kernel $K(u) = 0.75(1-u^2)_+$ is used for both examples.

5.1.1.1 Model I

We suppose

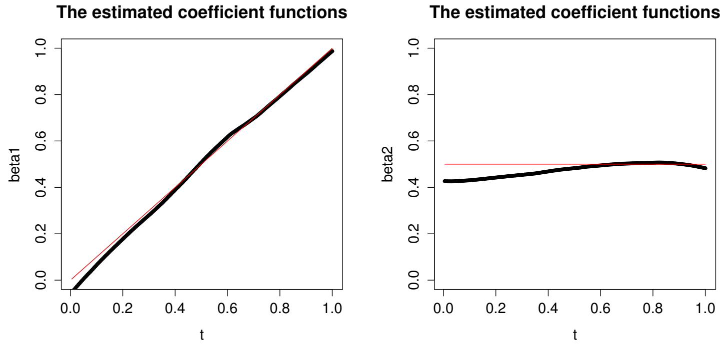
$$\lambda(t, x_1, x_2) = \exp(\beta_1(t)x_1 + \beta_2(t)x_2)$$

where $\lambda_0(t) = 1$ and $\beta_1(t) = t$ and $\beta_2(t) = 1/2$. The covariates are uniformly and normally distributed, respectively, i.e. $X_1 \sim \mathsf{U}(-1,1)$ and $X_2 \sim \mathsf{N}(0,1)$ with sample size $n = 600$. We consider the bandwidth $h = 0.6, 0.8$ and censoring rate 30%. Plots of the estimation at 200 grid points with $t = 0.005k, k = 1, ..., 200$ are shown in Figure 5.1 and 5.2.
The estimates of $\beta_1(t) = t$ and $\beta_2(t) = 1/2$ are plotted in (a) and (b), respectively. The red line is the true function. As can be seen in Figure 5.1 and 5.2, the estimates are close to the true functions and do not change very much if we take different bandwidth.



(a) $\widehat{\beta}_1$ (b) $\widehat{\beta}_2$

Figure 5.1: Estimation of $\beta_1(t) = t$ and $\beta_2(t) = \frac{1}{2}$ for Model I with $n = 600$, $h = 0.6$ and 30% censoring. The red lines are the true functions.



(a) $\widehat{\beta}_1$ (b) $\widehat{\beta}_2$

Figure 5.2: Estimation of $\beta_1(t) = t$ and $\beta_2(t) = \frac{1}{2}$ for Model I with $n = 600$, $h = 0.8$ and 30% censoring. The red lines are the true functions.

5.1.1.2 Model II

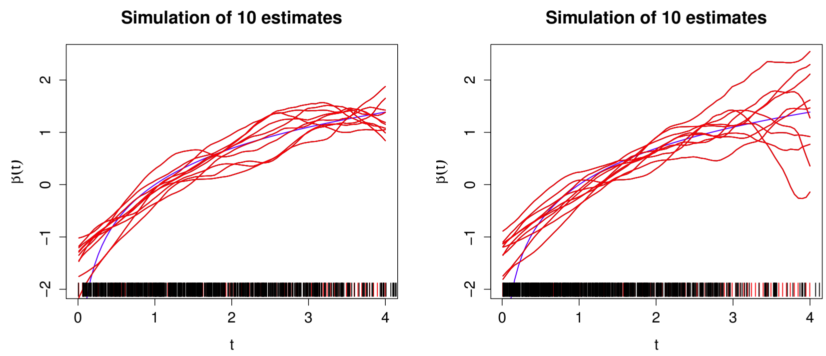
We consider

$$\lambda(t|x) = \lambda_0(t)\exp(\beta(t)x).$$

where $\lambda_0(t) = 1/2$, $\beta(t) = \log(t)$ and covariate X is generated from uniform distribution $\mathsf{U}(-1, 1)$.
For this simulation, we select sample sizes $n = 800$, band widths $h = 0.75$ and the two different censoring patterns 0% and 30%. The simulation of 10 estimates based on 80 grid points are displayed by Figure 5.3. The blue line is the true function. The results show that the estimates are quite good and they are quite close to the true function of the parameter except for $t > 3$ because there are less observations–marked at the bottom of the graph. These results on 0% censoring data are generally more reliable and stable than the results on 30% censoring data.

5.2 Score test based on the local partial likelihood approach

In this section the problem of testing whether the components of the coefficient function $\boldsymbol{\beta}_0(\cdot)$ have a prespecified parametric form is considered. Test procedures for such goodness-of-fit problem were studied by several



(a) 0% censoring (b) 30% censoring

Figure 5.3: Estimation of $\beta(t) = \log(t)$ for Model II with $n = 800$, $h = 0.75$ and 30% censoring. The blue lines are the true functions.

authors. Since these procedures are proposed for single component and not for the p-dimensional vector $\boldsymbol{\beta}_0$, without loss of generality let us consider the case $p = 1$. In Section (3.4) the results of Martinussen and Scheike (2000) were already mentioned. They proposed Cramer-von Mises and Kolmogorov type tests based on the nonparametric estimator for the cumulative coefficient $B_0(t) = \int_0^t \beta_0(s)ds$ for testing whether $\beta_0 \equiv 0$ or $\beta_0(t) \equiv \beta_0$ for a constant β_0.

Kauermann and Berger (2003) considered local constant estimators for $\beta_0(\cdot)$ and proposed the log partial likelihood ratio statistic to test whether $\beta_0(\cdot)$ is constant. They did not derive the (limit) distribution of the test statistic under the null hypothesis but applied a bootstrap techniques in order to obtain the reference distribution.

In the paper of Tian et al. (2005) confidence bands for β_0 based on the so-called strong approximation method along with a resampling procedure over a properly chosen time interval are derived. This interval can use to check $\beta_0(\cdot)$.

5.2.1 Distribution of the quadratic form of the score vector

We consider the test problem

$$\mathcal{H}: \beta_0(\cdot) \in B_{\text{par}} = \{\beta(\cdot, \vartheta),\ \vartheta \in \Theta \subseteq \mathbb{R}^k\} \qquad \mathcal{K}: \beta_0(\cdot) \notin B_{\text{par}}.$$

The most important special case of this null hypothesis is that $\beta_0(\cdot)$ is constant, that is, that the classical Cox proportional hazard model is true.
The test procedure will be based on the the local partial score function. Considering this score function at a finite number of different points we will show that the corresponding quadratic form converges in distribution to a χ^2-distribution. The advantage of this test procedure is that the computation of estimator is not required.
We will use the counting process approach introduced in Section 3.3.1 to show the results concerning inference procedures.
In Section 5.1 the log local constant partial likelihood function was defined as in (5.7).
The nonparametric local constant estimator of $\beta_0(\cdot)$ at the grid point t is the maximum of (5.7), for $p = 1$ the formula simplifies.

$$S_{kn}(b,u) \quad = \quad \frac{1}{n}\sum_{j=1}^{n} Y_j(u)\exp(bX_j)X_j^k \qquad k = 0,1,2$$

and with

$$E_n(b,t) = \frac{S_{1n}(b,u)}{S_{0n}(b,u)}$$

we can write the log local constant partial likelihood function (5.7)

$$\ell_t(b) \quad = \quad \sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\Big[bX_i - \log\big\{nS_{0n}(b,u)\big\}\Big]dN_i(u),$$

and the score function (5.4) as:

$$U_n(b,t) = \sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\Big[X_i - E_n(b,u)\Big]dN_i(u).$$

In the short survey in Chapter 3 it is mentioned that Cai and Sun (2003) proved the consistency of the local linear partial likelihood estimator $\widehat{\boldsymbol{\beta}}_n$ and its asymptotic normality at a fixed point. The proof of this limit statement is based on the asymptotic normality of the score function. Here we consider the vector of the score function at distinct grid point $t_1,\ldots,t_d$.
For $\boldsymbol{t} = (t_1,\ldots,t_d)$, we define

$$\boldsymbol{\mathcal{U}_n}(\beta,\boldsymbol{t}) = (U_n(\beta(t_1),t_1), U_n(\beta(t_2),t_2),\ldots,U_n(\beta(t_d),t_d))^T.$$

As an extension of the limit theorem for $U_n(\beta, t)$ at a fixed point t, we show that the distribution of $\boldsymbol{\mathcal{U}_n}(\beta_0, \boldsymbol{t})$ tends to a multivariate normal distribution with zero expectation and a covariance matrix $\mathcal{S}(\beta_0, \boldsymbol{t})$.
Then, using standard methods it follows that the corresponding quadratic form $\boldsymbol{\mathcal{U}_n}^T \mathcal{S}^{-1} \boldsymbol{\mathcal{U}_n}$ converges to a χ^2-distribution. This limit statement is the basis of the test procedure.
To formulate the multivariate limit theorem and the consequences for the test procedure we make use of the following assumptions. These assumptions include the assumptions formulated in the previous chapters, in addition conditions on the smoothness of the coefficient function, conditions on the kernel and the bandwidth and on the convergence rate are supposed.

A1 The coefficient function β_0 is twice continuously differentiable on $[0, \tau]$.

A2 The baseline function λ_0 is twice continuously differentiable on $[0, \tau]$.

B1 There exists a compact set $\mathcal{B}$ in $\mathbb{R}$ that includes a neighborhood of $\beta_0(t)$ for $t \in [0, \tau]$. Further, $s_j(\beta, t) = \mathsf{E} S_{jn}(\beta, t)$ exist for $j = 0, 1, 2$ and

$$|S_{jn}(\beta, t) - s_j(\beta, t)| = O_{\mathsf{P}}\left(n^{-1/2}\right) \quad \text{uniformly in } (\beta, t) \in \mathcal{B} \times [0, \tau]$$

B2 The functions s_j, $j = 0, 1, 2$, and their partial derivatives with respect to β are continuous in $\mathcal{B} \times [0, \tau]$.

B3 The functions $s_j(\beta_0(\cdot), \cdot)$ and $s_j(\beta, \cdot)$ for $j = 0, 1$ are twice differentiable with respect to $t \in [0, \tau]$.

B4 The function s_2 is bounded and s_0 is bounded away from zero.

C1 The function K is a symmetric density with bounded support, say $[-1, 1]$.

C2 The bandwidth sequence satisfies $h = h_n$

$$h_n \to 0 \quad \text{and} \quad nh_n^{1/5} \to 0 \quad \text{and} \quad nh \to \infty.$$

The asymptotic variance of the vector is characterized by the function

$$v(\beta, t) = \frac{s_2(\beta, t)}{s_0(\beta, t)} - e(\beta, t)^2 \qquad \text{with} \quad e(\beta, t) = \frac{s_1(\beta, t)}{s_0(\beta, t)}.$$

Now, let us formulate the theorem stating the asymptotic normality.

Theorem 5.1. *Suppose that the assumptions* A1, A2, B1 − B4, C1, C2 *are satisfied. If* $v(\beta_0(t_j), t_j) > 0$ *for all* $j = 1, \ldots, d$. *Then*

$$n^{-1/2}h^{1/2}\ \boldsymbol{\mathcal{U}_n}(\beta_0, \boldsymbol{t}) \xrightarrow{\mathrm{D}} \mathsf{N}_d(0, \mathcal{S}(\beta_0, \boldsymbol{t})),$$

where

$$\mathcal{S}(\beta_0, \boldsymbol{t}) = \mathrm{diag}(\sigma^2(\beta_0, t_1), \ldots, \sigma^2(\beta_0, t_d))$$

and

$$\sigma^2(\beta_0, t_j) = \kappa^2\, v(\beta_0(t_j), t_j) s_0(\beta_0(t_j), t_j)\lambda_0(t_j)$$

with $\kappa^2 = \int K^2(u)du$.

Proof. The proof consists of three steps. In the first step we apply the martingale decomposition; roughly speaking it is shown that under the formulated smoothness conditions on β_0 and under C2 it is enough to prove the asymptotic normality for the stochastic part of the score function. Then we follow the usual line–deriving the predictable variation process of the approximation process we obtain as limit the variance of the limiting process. Finally, the proof is completed by verifying the Lindeberg condition.

Each component of the standardized score vector $n^{-1/2}h^{1/2}\boldsymbol{\mathcal{U}_n}(\beta_0, \boldsymbol{t})$ can be decomposed as follows

$$\begin{aligned}
&n^{-1/2}h^{1/2}U_n(\beta_0(t), t)\\
&= n^{-1/2}h^{1/2}\sum_{i=1}^{n}\int_0^\tau K_h(s-t)\Big[X_i - E_n(\beta_0(t), s)\Big]dM_i(s)\\
&\quad + n^{-1/2}h^{1/2}\sum_{i=1}^{n}\int_0^\tau K_h(s-t)\Big[X_i - E_n(\beta_0(t), s)\Big]Y_i(s)\lambda(s, X_i)ds\\
&= n^{-1/2}h^{1/2}\sum_{i=1}^{n}\int_0^\tau K_h(s-t)\Big[X_i - E_n(\beta_0(t), s)\Big]dM_i(s) &(5.8)\\
&\quad + n^{-1/2}h^{1/2}\sum_{i=1}^{n}\int_0^\tau K_h(s-t)\Big[X_i - E_n(\beta_0(s), s)\Big]Y_i(s)\lambda(s, X_i)ds &(5.9)\\
&\quad + n^{-1/2}h^{1/2}\sum_{i=1}^{n}\int_0^\tau K_h(s-t)\Big[E_n(\beta_0(s), s) - E_n(\beta_0(t), s)\Big]Y_i(s)\lambda(s, X_i)ds. &(5.10)
\end{aligned}$$

The term (5.9) is equal to zero. We will show that the third term (5.10) can be neglected, so that it is enough to prove the asymptotic normality of the vector with components (5.8).
Consider

$$\begin{aligned} & R_n(t) \\ = \ & n^{-1/2}h^{1/2}\sum_{i=1}^{n}\int_0^{\tau} K_h(s-t)\Big[E_n(\beta_0(s),s)-E_n(\beta_0(t),s)\Big]Y_i(s)\lambda(s,X_i)ds \\ = \ & (nh)^{1/2}\int_0^{\tau} K_h(s-t)\Big[E_n(\beta_0(s),s)-E_n(\beta(t),s)\Big]S_{0n}(\beta_0(s),s)\lambda_0(s)ds. \end{aligned}$$

By the consistency of the functions S_{jn} (uniformly with respect to β and t) we can replace the term $[E_n(\beta_0(s),s)-E_n(\beta_0(t),s)]S_{0n}(\beta_0(s),s)$ by their limits $[e(\beta_0(s),s)-e(\beta_0(t),s)]s_0(\beta_0(s),s)$. Furthermore, for

$$R_n^{(1)}(t) \ = \ (nh)^{1/2}\int_0^{\tau} K_h(s-t)\Big[e(\beta_0(s),s)-e(\beta_0(t),s)\Big]s_0(\beta_0(s),s)\lambda_0(s)ds$$

we have

$$\sup_t |R_n(t) \ - \ R_n^{(1)}(t)| = O_{\mathsf{P}}\Big(h^{1/2}\Big).$$

Moreover, since the function β is twice continuously differentiable and K is a symmetric kernel it follows by Taylor expansion

$$\int_0^{\tau} K_h(s-t)\Big[e(\beta_0(s),s)-e(\beta_0(t),s)\Big]s_0(\beta_0(s),s)\lambda_0(s)ds = O_{\mathsf{P}}\Big(h^2\Big),$$

therefore

$$R_n^{(1)}(t) = O_{\mathsf{P}}\Big(\sqrt{nh^5}\Big).$$

Thus, with

$$U_n^{(1)}(\beta_0(t),t) = \sum_{i=1}^{n}\int_0^{\tau} K_h(s-t)\Big[X_i - E_n(\beta_0(t),s)\Big]dM_i(s)$$

we have by the condition on the bandwidth

$$n^{-1/2}h^{1/2}U_n(\beta_0(t),t) = n^{-1/2}h^{1/2}U_n^{(1)}(\beta_0(t),t) + o_{\mathsf{P}}(1). \qquad (5.11)$$

By (5.11) it is enough to prove that the vector

$$
\begin{aligned}
& n^{-1/2}h^{1/2}\boldsymbol{\mathcal{U}_n}^{(1)}(\beta_0,\boldsymbol{t}) \\
= \; & n^{-1/2}h^{1/2}(U_n^{(1)}(\beta_0(t_1),t_1),U_n^{(1)}(\beta_0(t_2),t_2),\ldots,U_n^{(1)}(\beta_0(t_k),t_k))^T
\end{aligned}
$$

is asymptotically normal. For this purpose we apply Rebolledo's Central Limit Theorem for local square integrable martingales in the form given in the Appendix of Andersen and Gill (1982):
For arbitrary j,m and s we consider the predictable variation process, which is denoted by $\langle\cdot\rangle$,

$$
\begin{aligned}
& \langle n^{-1/2}h^{1/2}U_n^{(1)}(\beta_0(t_j),t_j),n^{-1/2}h^{1/2}U_n^{(1)}(\beta_0(t_m),t_m)\rangle(s) \\
= \; & n^{-1}h\int_0^s\sum_{i=1}^n K_h(u-t_j)K_h(u-t_m)\;\times \\
& \times\Big[X_i-E_n(\beta_0(t_j),u)\Big]\Big[X_i-E_n(\beta_0(t_m),u)\Big]\alpha_i(u)du.
\end{aligned}
$$

Set $W_{ni}(t,u)=X_i-E_n(\beta_0(t),u)$. Then we have for $t_j\neq t_m$

$$
\begin{aligned}
& \langle n^{-1/2}h^{1/2}U_n^{(1)}(\beta_0(t_j),t_j),n^{-1/2}h^{1/2}U_n^{(1)}(\beta_0(t_m),t_m)\rangle(s) \\
= \; & n^{-1}\sum_{i=1}^n\int_0^s K(u)K\left(\frac{t_j-t_m}{h}+u\right)W_{ni}(t_j,t_j+hu)W_{ni}(t_m,t_j+hu)\;\times \\
& \times Y_i(t_j+hu)\exp(X_i\beta_0(t_j+hu))\lambda_0(t_j+hu)du \\
\xrightarrow{\text{P}} \; & 0\;\;\text{as } h\to 0.
\end{aligned}
$$

For $t_j=t_m$ we obtain

$$
\begin{aligned}
& \langle n^{-1/2}h^{1/2}U_n^{(1)}(\beta_0(t_j),t_j),n^{-1/2}h^{1/2}U_n^{(1)}(\beta_0(t_j),t_j)\rangle(s) \\
= \; & n^{-1}\sum_{i=1}^n\int_0^s K^2(u)\,W_{ni}^2(t_j,t_j+hu)Y_i(t_j+hu)\;\times \\
& \times\exp(X_i\beta_0(t_j+hu))\lambda_0(t_j+hu)du \\
= \; & \kappa^2\left(\frac{S_{2n}(\beta_0(t_j),t_j)}{S_{0n}(\beta_0(t_j),t_j)}-E_n^2(\beta_0(t_j),t_j)\right)S_{0n}(\beta_0(t_j),t_j)\lambda_0(t_j)+o_{\mathsf{P}}(1) \\
= \; & \kappa^2V_n(\beta_0(t_j),t_j)S_{0n}(\beta_0(t_j),t_j)\lambda_0(t_j)+o_{\mathsf{P}}(1) \\
= \; & \kappa^2v(\beta_0(t_j),t_j)s_0(\beta_0(t_j),t_j)\lambda_0(t_j)+o_{\mathsf{P}}(1)
\end{aligned}
$$

where

$$V_n(\beta_0(t_j),t_j) = \frac{S_{2n}(\beta_0(t_j),t_j)}{S_{0n}(\beta_0(t_j),t_j)} - \left(\frac{S_{1n}(\beta_0(t_j),t_j)}{S_{0n}(\beta_0(t_j),t_j)}\right)^2.$$

Now, it remains to check the Lindeberg condition: From the smoothness conditions formulated above it follows that for all t_j and all $\varepsilon > 0$

$$n^{-1}h\sum_{i=1}^{n}\int_0^{\tau} K_h^2(u-t_j)W_{ni}^2(t_j,u)\mathbb{1}(\sqrt{h/n}K_h(u-t_j)|W_{ni}(u)| > \varepsilon)\ \alpha_i(u)du$$

tends in probability to zero. The proof of the theorem is complete. □

This is the illustration of the asymptotic normality of the score function. Figure 5.4 shows the simulated realization of the score function at two grid points together with the contour lines of the approximating normal distribution.

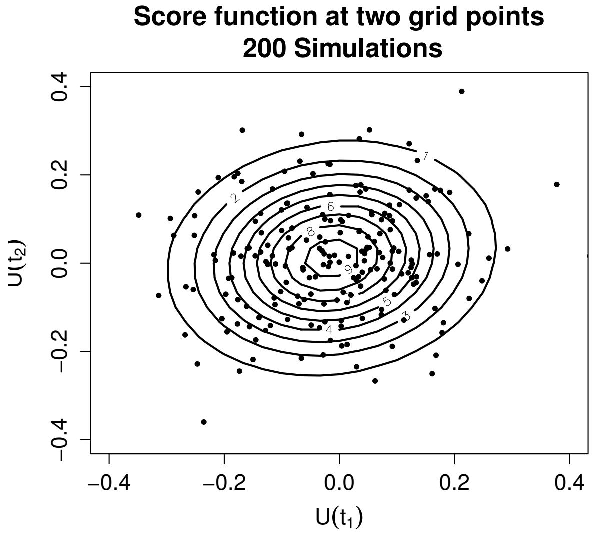


Figure 5.4: The score function at 2 grid points, 200 simulations .

Consider the weighted quadratic form

$$\mathcal{T}_n(\beta_0) = \boldsymbol{\mathcal{U}_n}(\beta_0,\boldsymbol{t})^T\mathcal{S}^{-1}(\beta_0,\boldsymbol{t})\ \boldsymbol{\mathcal{U}_n}(\beta_0,\boldsymbol{t}) = \sum_{j=1}^{d} U_n^2(\beta_0(t_j),t_j)\sigma^{-2}(\beta_0,t_j).$$

From the asymptotic normality of the vector $\boldsymbol{\mathcal{U}_n}(\beta_0, \boldsymbol{t})$ we obtain the following corollary:

Corollary 5.2. *Under the assumptions of Theorem 5.1*

$$n^{-1}h\mathcal{T}_n(\beta_0) \xrightarrow{\text{D}} \chi^2_d.$$

Proof. The proof is straightforward. Since the vector $n^{-1/2}h^{1/2}\boldsymbol{\mathcal{U}_n}$ is asymptotically normal, and the limiting variance matrix $\mathcal{S}$ is positive definite the quadratic form $n^{-1}h\ \boldsymbol{\mathcal{U}_n}^T\mathcal{S}^{-1}\boldsymbol{\mathcal{U}_n}$ converges in distribution to a χ^2-distribution with d degrees of freedom. □

The variance matrix $\mathcal{S}$ is unknown. It depends on the unknown limits of the sums S_{kn}, on β_0 and on λ_0. A consistent estimator of $\mathcal{S}(\beta_0, \boldsymbol{t})$ is given by

$$\widehat{\mathcal{S}}_n(\widehat{\beta}_n, \boldsymbol{t}) = \operatorname{diag}(\widehat{\sigma}^2_n(\widehat{\beta}_n, t_1), \ldots, \widehat{\sigma}^2_n(\widehat{\beta}_n, t_d))$$

with

$$\widehat{\sigma}^2_n(\beta, t_j) = \kappa^2\,\frac{1}{n}\sum_{i=1}^{n}\int K_h(u - t_j)V_n(\beta, u)dN_i(u)$$

where

$$V_n(\beta, t) = \frac{S_{2n}(\beta, t)}{S_{0n}(\beta, t)} - \left(\frac{S_{1n}(\beta, t)}{S_{0n}(\beta, t)}\right)^2.$$

and $\widehat{\beta}_n(t)$ is the estimator of $\beta_0(t)$.
In the following corollary we show that the limit statement remain true if $\mathcal{S}$ is replaced be the estimates $\widehat{\mathcal{S}}_n(\widehat{\beta}_n, \boldsymbol{t})$.

Corollary 5.3. *Under the assumptions of Theorem 5.1*

$$n^{-1}h\tilde{\mathcal{T}}_n(\beta_0) \xrightarrow{\text{D}} \chi^2_d$$

where

$$\tilde{\mathcal{T}}_n(\beta_0) = \boldsymbol{\mathcal{U}_n}(\beta_0, \boldsymbol{t})^T\widehat{\mathcal{S}}_n^{-1}(\widehat{\beta}_n, \boldsymbol{t})\boldsymbol{\mathcal{U}_n}(\beta_0, \boldsymbol{t}).$$

Proof. To prove Corollary 5.3 it is enough to verify the consistency of the variance estimator $\widehat{\mathcal{S}}_n$. For this purpose consider for an arbitrary component of the diagonal matrix the term

$$v(\beta_0(t), t)s_0(\beta_0(t), t))\lambda_0(t)$$

and its estimator

$$\int K_h(u-t)V_n(\widehat{\beta}_n(t),u)d\overline{N}_n(u) \qquad \text{with} \quad \overline{N}_n(u) = \frac{1}{n}\sum_{i=1}^{n} N_i(u).$$

We have

$$\begin{aligned}
& \Big|\int K_h(u-t)V_n(\widehat{\beta}_n(t),u)d\overline{N}_n(u) - v(\beta_0(t),t)s_0(\beta_0(t),t))\lambda_0(t)\Big| \\
\leq\ & \Big|\int K_h(u-t)(V_n(\widehat{\beta}_n(t),u) - v(\widehat{\beta}_n(t),u))d\overline{N}_n(u)\Big| && (5.12)\\
& +\Big|\int K_h(u-t)(v(\widehat{\beta}_n(t),u) - v(\beta_0(t),u))d\overline{N}_n(u)\Big| && (5.13)\\
& +\Big|\int K_h(u-t)v(\beta_0(t),u)(d\overline{N}_n(u) - S_{0n}(\beta_0(t),u)\lambda_0(u)du)\Big| && (5.14)\\
& +\Big|\int K_h(u-t)v(\beta_0(t),u)(S_{0n}(\beta_0(t),u) - s_0(\beta_0(t),u))\lambda_0(u)du)\Big| && (5.15)\\
& +\Big|\int K_h(u-t)v(\beta_0(t),u)s_0(\beta_0(t),u)\lambda_0(u)du \\
& -v(\beta_0(t),t)s_0(\beta_0(t),t))\lambda_0(t)\Big|. && (5.16)
\end{aligned}$$

The uniform consistency of the functions S_{jn} and the boundedness of s_0 imply that the terms (5.12) and (5.15) tend to zero (in probability). For term (5.13) the same statement follows from the consistency of $\widehat{\beta}_n$, and (5.16) tends to zero because of the smoothness of the functions s_j. From the inequality of Lenglart for local martingales it follows that the term (5.14) converges also to zero. □

5.2.2 Formulation of the test procedure

Consider now the hypothesis that the coefficient function $\beta_0(\cdot)$ has a parametric form, say $\beta_0(\cdot) = \beta(\cdot;\vartheta_0)$ for some unknown parameter ϑ_0. Thus, the test problem is

$$\mathcal{H} : \beta_0(\cdot) \in B_{\text{par}} = \{\beta(\cdot,\vartheta),\ \vartheta \in \Theta \subseteq \mathbb{R}^k\} \qquad \mathcal{K} : \beta_0(\cdot) \notin B_{\text{par}}.$$

To estimate the parameter ϑ under $\mathcal{H}$ we use the partial likelihood method in the hypothetical model B_{par}

$$\lambda_i(t) = \lambda_0(t)\exp(\beta(t,\vartheta)x_i).$$

The partial likelihood function of ϑ is denoted by $\tilde{\ell}$

$$\tilde{\ell}(\vartheta) = \sum_{i=1}^{n} \int_0^{\tau} \Big[\beta(s,\vartheta)X_i - \log\Big\{\sum_{j=1}^{n} Y_j(s)\exp(\beta(s,\vartheta)X_j)\Big\}\Big] dN_i(s).$$

Let C_n be the corresponding score vector, i.e., its component C_{nr} $r = 1, \dots, k$ is

$$C_{nr}(\vartheta) = \frac{\partial \tilde{\ell}(\vartheta)}{\partial \vartheta_r} = \sum_{i=1}^{n} \int_0^{\tau} \Big[X_i - E_n(\beta(s,\vartheta), s)\Big] \dot{\beta}_r(s,\vartheta) dN_i(s)$$

where $\dot{\beta}_r(t,\vartheta)$ is the partial derivative of $\beta(t,\vartheta)$ with respect to ϑ_r. The estimator $\widehat{\vartheta}_n$ is the solution of the system of equations

$$C_{nr}(\vartheta) = 0 \quad r = 1, \dots, k. \tag{5.17}$$

Suppose that the hypothesis $\mathcal{H}$ is true, i.e. there exist a ϑ_0 such that $\beta_0(\cdot) = \beta(\cdot, \vartheta_0)$.
If the estimator $\widehat{\vartheta}_n$ is $\sqrt{n}$-consistent, we can apply $\widehat{\vartheta}_n$ in the test procedure. To verify $\sqrt{n}$-consistency we will show that $\widehat{\vartheta}$ is asymptotically normal. The proof is based on the following steps: If the partial likelihood function $\tilde{\ell}$ is strictly concave, then the solution to (5.17) is unique. The consistency follows by showing, that the partial likelihood function converges to a concave function with a unique maximum at the underlying ϑ_0. To obtain the rate of convergence we consider the score function C_n as a local martingale and prove that the matrix of the minus second derivatives converges to a positive definite matrix.

Theorem 5.4. *Suppose that the hypothetical functions in B_{par} have continuous partial derivatives of second order with respect to ϑ. Further assume that the partial likelihood function is strictly concave. Let the assumptions A2 and B1-B4 be satisfied. Then*

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) = O_p(1). \tag{5.18}$$

Proof. Define

$$
\begin{aligned}
Q_n &= n^{-1}(\tilde{\ell}(\vartheta) - \tilde{\ell}(\vartheta_0)) \\
=& n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\Big\{\beta(u,\vartheta)-\beta(u,\vartheta_0)\Big\}X_i - \log\left\{\frac{S_{0n}(\beta(u,\vartheta),u)}{S_{0n}(\beta(u,\vartheta_0),u)}\right\}\right]dN_i(u) \\
=& n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\Big\{\beta(u,\vartheta)-\beta(u,\vartheta_0)\Big\}X_i - \log\left\{\frac{S_{0n}(\beta(u,\vartheta),u)}{S_{0n}(\beta(u,\vartheta_0),u)}\right\}\right]dM_i(u) \\
&+ n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\Big\{\beta(u,\vartheta)-\beta(u,\vartheta_0)\Big\}X_i - \log\left\{\frac{S_{0n}(\beta(u,\vartheta),u)}{S_{0n}(\beta(u,\vartheta_0),u)}\right\}\right]dA_i(u) \\
=& Q_{n1} + Q_{n2}.
\end{aligned}
$$

The second summand is equal to

$$
\begin{aligned}
Q_{n2} \;=\; & \int_0^{\tau}\left[\Big\{\beta(u,\vartheta)-\beta(u,\vartheta_0)\Big\}S_{1n}(\beta(u,\vartheta_0),u)\right. \\
& \left. -\log\left\{\frac{S_{0n}(\beta(u,\vartheta),u)}{S_{0n}(\beta(u,\vartheta_0),u)}\right\}S_{0n}(\beta(u,\vartheta_0),u)\right]\lambda_0(u)du.
\end{aligned}
$$

Q_{n1} is a local square integrable martingale with the predictable variation process

$$
\begin{aligned}
&\langle Q_{n1}(\vartheta), Q_{n1}(\vartheta)\rangle \\
=& n^{-2}\sum_{i=1}^{n}\int_0^{\tau}\left[\Big\{\beta(u,\vartheta)-\beta(u,\vartheta_0)\Big\}X_i - \log\left\{\frac{S_{0n}(\beta(u,\vartheta),u)}{S_{0n}(\beta(u,\vartheta_0),u)}\right\}\right]^2 dA_i(u) \\
=& n^{-1}\int\left[\Big\{\beta(u,\vartheta)-\beta(u,\vartheta_0)\Big\}^2 S_{2n}(\beta(u,\vartheta_0),u)\right. \\
&+ \left[\log\left\{\frac{S_{0n}(\beta(u,\vartheta),u)}{S_{0n}(\beta(u,\vartheta_0),u)}\right\}\right]^2 S_{0n}(\beta(u,\vartheta_0),u) \\
&\left. - 2\big[\beta(u,\vartheta)-\beta(u,\vartheta_0)\big]\left[\log\left\{\frac{S_{0n}(\beta(u,\vartheta),u)}{S_{0n}(\beta(u,\vartheta_0),u)}\right\}\right]S_{1n}(\beta(u,\vartheta_0),u)\right]\lambda_0(u)du.
\end{aligned}
$$

By the conditions of the theorem it follows that $\langle Q_{n1}(\vartheta), Q_{n1}(\vartheta)\rangle = O_{\mathsf{P}}(n^{-1})$, thus Q_n converges to the same limit as Q_{n2}. The smoothness conditions on the sums S_{kn} imply

$$\begin{aligned} Q_{n2}(\vartheta) \quad &\xrightarrow{\mathrm{P}} \quad \int_0^\tau \Bigg[\Big\{ \beta(u,\vartheta) - \beta(u,\vartheta_0) \Big\} s_1(\beta(u,\vartheta_0),u) \\ &\qquad - \log\left\{ \frac{s_0(\beta(u,\vartheta),u)}{s_0(\beta(u,\vartheta_0),u)} \right\} s_0(\beta(u,\vartheta_0),u) \Bigg] \lambda_0(u)du \\ &=: \quad Q_2(\vartheta). \end{aligned}$$

Let us compute the derivatives of the limit Q_2. (Here we can change integration and differentiation because of the boundedness of the integrand.)

$$\begin{aligned} \frac{\partial Q_2(\vartheta)}{\partial \vartheta} &= \int_0^\tau \left[s_1(\beta(u,\vartheta_0),u) - s_1(\beta(u,\vartheta),u) \frac{s_0(\beta(u,\vartheta_0),u)}{s_0(\beta(u,\vartheta),u)} \right] \dot\beta(u,\vartheta)\lambda_0(u)du \\ &= \int_0^\tau \Big(e(\beta(u,\vartheta_0),u) - e(\beta(u,\vartheta),u) \Big) s_0(\beta(u,\vartheta_0),u) \dot\beta(u,\vartheta)\lambda_0(u)du. \end{aligned}$$

The second derivative is

$$\begin{aligned} &- \int_0^\tau v(\beta(u,\vartheta),u)\dot\beta(u,\vartheta)^{\otimes 2} s_0(\beta(u,\vartheta_0),u)\lambda_0(u)du \\ &+ \int_0^\tau \Big(e(\beta(u,\vartheta_0),u) - e(\beta(u,\vartheta),u) \Big) s_0(\beta(u,\vartheta),u)\ddot\beta(u,\vartheta)\lambda_0(u)du. \end{aligned}$$

The first derivative is zero at $\vartheta = \vartheta_0$, the second is minus a positive definite matrix at $\vartheta = \vartheta_0$. Thus, the limiting function of $\tilde l_n$ has a unique maximum at $\vartheta = \vartheta_0$. It follows that $\widehat{\vartheta_n}$, the unique maximizer of $\tilde l_n$ converges to ϑ_0. □

Based on Theorem 5.4, it follows that under $\mathcal{H}$

$$n^{-1} h \widehat{\mathcal{T}}_n \xrightarrow{\mathrm{D}} \chi^2_d \tag{5.19}$$

where

$$\widehat{\mathcal{T}}_n = \widehat{\boldsymbol{\mathcal{U}}}_n(\beta_{\widehat{\vartheta}_n}, \boldsymbol{t})^T \, \widehat{\mathcal{S}}_n^{-1}(\beta_{\widehat{\vartheta}_n}, \boldsymbol{t}) \widehat{\boldsymbol{\mathcal{U}}}_n(\beta_{\widehat{\vartheta}_n}, \boldsymbol{t}) \tag{5.20}$$

with

$$\widehat{\boldsymbol{\mathcal{U}}}_n(\beta_{\widehat{\vartheta}_n}, \boldsymbol{t}) = (U_n(\beta(t_1,\widehat{\vartheta}_n),t_1), U_n(\beta(t_2,\widehat{\vartheta}_n),t_2), \ldots, U_n(\beta(t_d,\widehat{\vartheta}_n),t_d))^T.$$

Limit statement (5.19) implies the following asymptotic test procedure. Reject $\mathcal{H}$, iff

$$n^{-1} h \widehat{\mathcal{T}}_n \geq \chi^2_{d;1-\alpha}. \tag{5.21}$$

5.3 Bootstrap version of the score test

In this section we consider the problem of testing whether the coefficient function is constant. As test statistic the score statistic is applied, however to determine the critical value or the p-value, respectively, we make use of bootstrapping.
The null hypothesis has the form

$$\mathcal{H} : \beta_0(\cdot) \equiv \vartheta \quad \text{for some constant } \vartheta. \tag{5.22}$$

The test statistic is given by

$$\widehat{\mathcal{T}}_n = \widehat{\boldsymbol{\mathcal{U}}}_n(\widehat{\vartheta}_n, \boldsymbol{t})^T \, \widehat{\mathcal{S}}_n^{-1}(\widehat{\vartheta}_n, \boldsymbol{t}) \widehat{\boldsymbol{\mathcal{U}}}_n(\widehat{\vartheta}_n, \boldsymbol{t})$$

where

$$\widehat{\boldsymbol{\mathcal{U}}}_n(\widehat{\vartheta}_n, \boldsymbol{t}) = (U_n(\widehat{\vartheta}_n, t_1), U_n(\widehat{\vartheta}_n, t_2), \ldots, U_n(\widehat{\vartheta}_n, t_d))^T$$

with

$$U_n(\widehat{\vartheta}_n, t_j) = n^{-1/2} h^{1/2} \sum_{i=1}^{n} \int_0^{\tau} K_h(s - t_j) \Big[X_i - E_n(\widehat{\vartheta}_n, s) \Big] dN_i(s); \; j = 1, \ldots, d$$

where $\widehat{\vartheta}_n$ is the maximum partial likelihood estimator in the hypothetical model.

The bootstrap algorithm which is used based on the procedure `censboot` function in the R software package suggested by Davison and Hinkley (1997). Let us describe this idea of this resampling procedure.

The aim is to generate data (T_i^*, Δ_i^*) with

$$T_i^* = \min(\tilde{T}_i^*, C_i^*)$$

and the censoring indicator $\Delta_i^* = \mathbb{1}(\tilde{T}_i^* \leq C_i^*)$ where C_i^* is the censoring time and $\tilde{T}_i^*$ describes the lifetime with the distribution under $\mathcal{H}$, i.e. $1 - H(\cdot|X_i)$ with $1 - H(\cdot|X_i) = (1 - F(\cdot|X_i))(1 - G(\cdot))$ and $F(t|X_i) = 1 - S_0(t)^{\exp(\vartheta X_i)}$ where S_0 is the baseline survival function.
To do this we estimate the conditional distribution of $\tilde{T}_i$ under $\mathcal{H}$.
Estimating ϑ in the classical Cox model by MPLE we obtain $\widehat{\vartheta}_n$; applying the Breslow estimator the cumulative hazard function Λ_0 is estimated by

$$\widehat{\Lambda}_{0n}(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \frac{1}{S_{0n}(u, \widehat{\vartheta}_n)} dN_i(u).$$

Then the baseline survival function is estimated by

$$\widehat{S}_{0n}(t) = \exp(-\widehat{\Lambda}_0(t))$$

and an estimator of $F(\cdot|x)$ is given by

$$\widehat{F}_n(t|x) = 1 - \widehat{S}_{0n}(t)^{\exp(\widehat{\vartheta}_n x)}.$$

Figure 5.5 shows examples of these estimates.

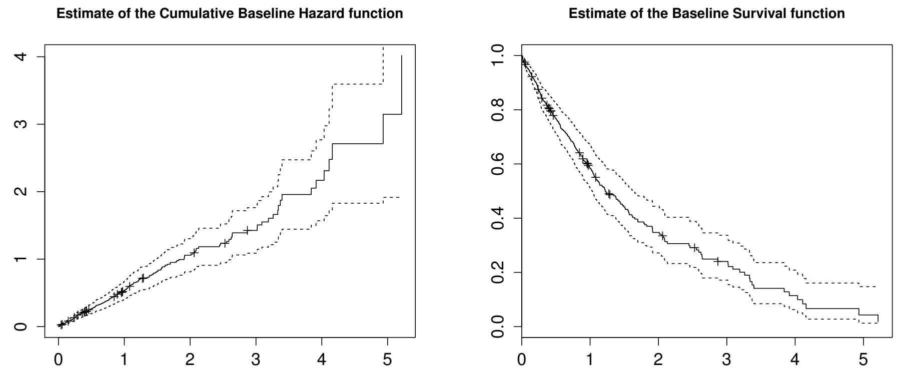


Figure 5.5: The estimated cumulative baseline hazard function and the estimated baseline survival function under $\mathcal{H}$.

The $\tilde{T}_i^*$ are generated according to $\widehat{F}_n(t|X_i)$.
An estimator for the distribution function of the censoring variables can be constructed by the Kaplan-Meier method, where instead of Δ_i's the observations $1 - \Delta_i$ are taken, then one defines

$$1 - \widehat{G}_n(t) = \prod_{i:T_i \le t} \Big(\frac{n-i}{n-i+1} \Big)^{1-\Delta_i}.$$

We will not generate C_i^* according to $\widehat{G}_n$, but we will apply the so-called conditional resampling to generate C_i^*.
The idea of this method can be explained as follows. Consider the distribution of C_i given T_i and Δ_i: The conditional distribution of C_i given T_i and $\Delta_i = 0$ is the one-point distribution at T_i. The conditional distribution of C_i given T_i and $\Delta_i = 1$ is given by

$$\mathsf{P}(C_i \le c | \Delta_i = 1, T_i) = \frac{G(c) - G(T_i^-)}{1 - G(T_i^-)}.$$

Davison and Hinkley (1997) proposed to set $C_i^* = T_i$ for $\Delta_i = 0$ and C_i^* is generated according to the distribution $\dfrac{\widehat{G}_n(\cdot) - \widehat{G}_n(T_i)}{1 - \widehat{G}_n(T_i)}$ for $\Delta_i = 1$.

Now let us formulate the bootstrap version of the score test for (5.22):

1. Construct the estimates $\widehat{\vartheta}_n$, $\widehat{\Lambda}_{0n}$, $\widehat{G}_n$ and $\widehat{F}_n(t|X_i)$ for $i = 1, \dots, n$;
2. Generate $\tilde{T}_1^*, \dots, \tilde{T}_n^*$ according to the distribution $\widehat{F}_n(t|X_i)$;
3. For $\Delta_i = 0$, set $C_i^* = T_i$ and for $\Delta_i = 1$, generate C_i^* from the distribution $\dfrac{\widehat{G}_n(\cdot) - \widehat{G}_n(T_i)}{1 - \widehat{G}_n(T_i)}$;
4. Define T_i^* by $T_i^* = \min(\tilde{T}_i^*, C_i^*)$ and set $\Delta_i^* = \mathbb{1}(\tilde{T}_i^* \leq C_i^*)$;
5. Calculate
$$\widehat{\mathcal{T}}_n^* = \widehat{\boldsymbol{\mathcal{U}}_{\boldsymbol{n}}^*}(\widehat{\vartheta}_n, \boldsymbol{t})^T \,\widehat{\mathcal{S}^*}_n^{-1} (\widehat{\vartheta}_n, \boldsymbol{t}) \,\widehat{\boldsymbol{\mathcal{U}}_{\boldsymbol{n}}^*}(\widehat{\vartheta}_n, \boldsymbol{t});$$
6. Repeat the steps 2 to 5 R times to obtain $\widehat{\mathcal{T}}_{n(1)}^*, \dots, \widehat{\mathcal{T}}_{n(R)}^*$.

The bootstrap p-value is given by

$$p_{boot} = \frac{\sum_{r=1}^{R} \mathbb{1}(\widehat{\mathcal{T}}_{n(r)}^* > \widehat{\mathcal{T}}_{n(0)}) + 1}{R + 1}$$

where $\widehat{\mathcal{T}}_{n(0)}$ is the value of test statistic with the original data.

5.3.1 Simulation Study

In this section, we investigate the behavior of the bootstrap score test under the null hypothesis and a special alternative hypothesis.

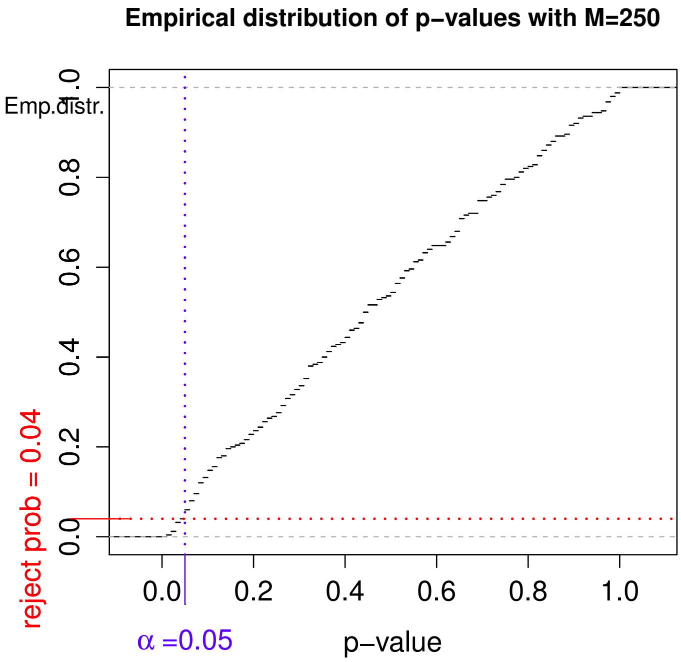
In each simulation $M = 250$ times the bootstrap test was carried out, $R = 99$ bootstrap replicates were computed. The sample size was $n = 400$. The bandwidth was taken to be 0.2 and the six grid points $t = 0.3, 0.4, 0.6, 0.8, 1, 1.2$ were used throughout.

In the first simulation study data according to $\beta_0(t) = 2$, $\lambda_0(t) = 1/2$, $X_i \sim \mathsf{U}(0, 1)$ and $C_i \sim \mathsf{Exp}(a)$, where a was selected in such a way that it occurs 30% censoring, were generated and the p-values were determined.

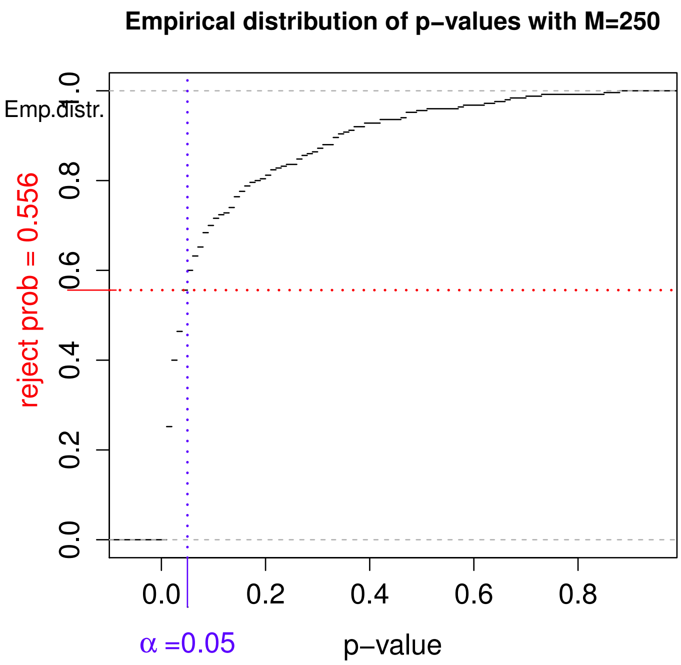
Figure 5.6(a) shows the p-values distribution. This figure shows that the empirical distribution of the 250 p-values is similar to a uniform distribution. This result corresponds to the fact that the p-values are uniformly distributed under the null hypothesis. Taking the significance level $\alpha = 0.05$ we see that 4% of the simulations would reject the null hypothesis.

In the next example data according to an alternative model were generated. The coefficient function is taken by $\beta_0(t) = \log(t)$, the underlying baseline distribution, the censoring distribution and the covariate distribution is taken as before, i.e. $\lambda_0(t) = 1/2$, $X \sim \mathsf{U}(0,1)$ and $C_i \sim \mathsf{Exp}(a)$, with 99 bootstrap replicates. The empirical distribution of the 250 p-values is given in Figure 5.6(b). Here we see that taking $\alpha = 0.05$ we have the estimated power (w.r.t. this alternative) is 55.6%.

The simulation study is only a first step. To compare the power, several different alternative setting should be considered, such as $\beta_0(t) = t^2$, $\beta_0(t) = 1 - t/50$ etc.
Moreover, topics for further investigations is to study the influence of choice of the bandwidth and the influence of the censoring on the performance on the test.



(a)



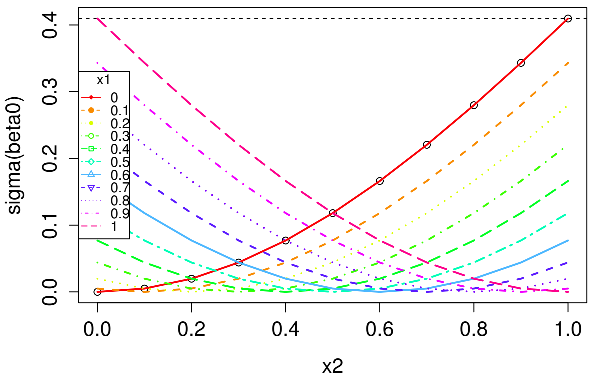
(b)

Figure 5.6: Empirical distribution of p-value for the score test in the first simulation, i.e. $\beta_0 = 2$ (left plot) and the second simulation, i.e. $\beta_0(t) = \log(t)$ (right plot)

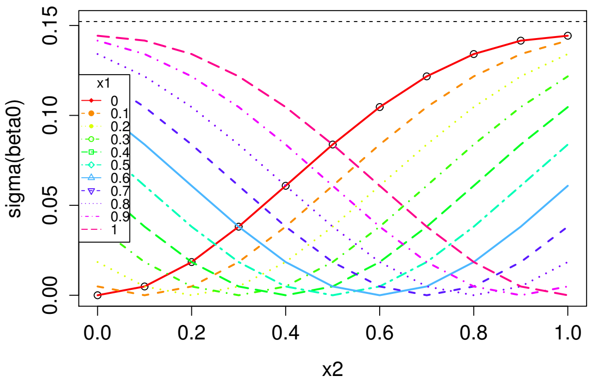
.

Appendix A

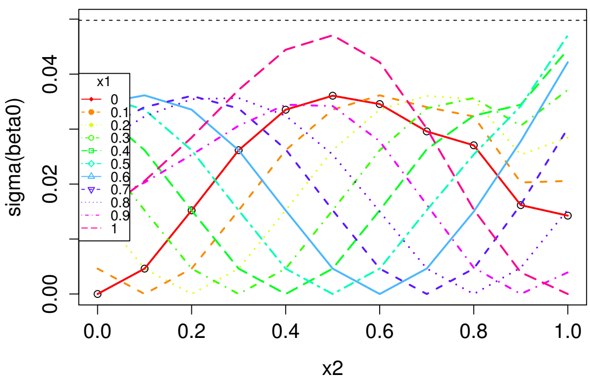
Figures and Tables for Chapter 4



(a) $\beta_0 = 1.0$

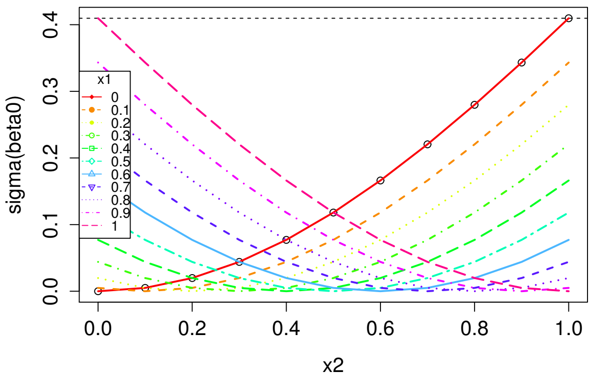


(b) $\beta_0 = 3.0$

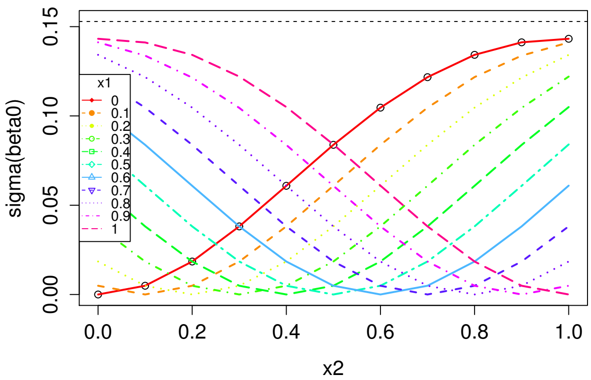


(c) $\beta_0 = 6.0$

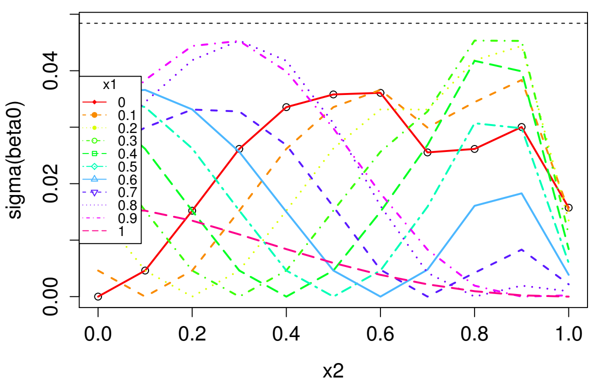
Figure A.1: The inverse of the asymptotic variance for the exponential model in dependence of the value of ξ_2 for several ξ_1. The values for $\lambda_0 = 1$ and $q_1 = 0.5$ are fixed, β_0 takes three different values.



(a) $\beta_0 = 1.0$

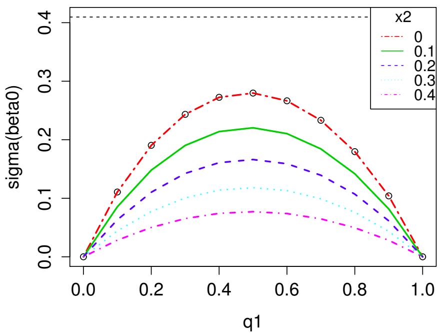


(b) $\beta_0 = 3.0$

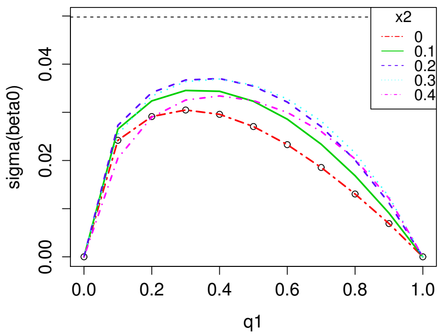


(c) $\beta_0 = 6.0$

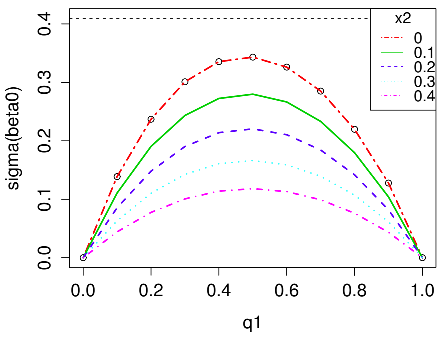
Figure A.2: The inverse of the asymptotic variance for the exponential model in dependence of the value of ξ_2 for several ξ_1. The values for $\lambda_0 = 3$ and $q_1 = 0.5$ are fixed, β_0 takes three different values.



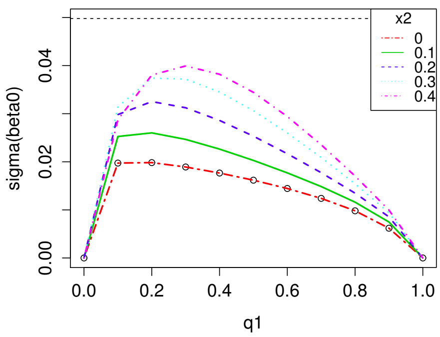
(a) $\beta_0 = 1.0, \xi_1 = 0.8$



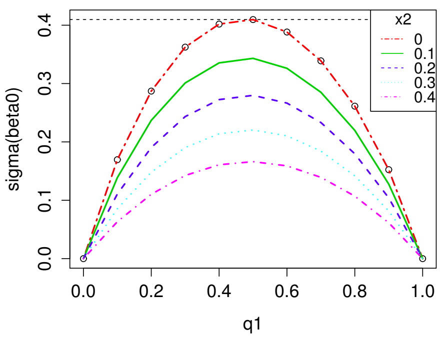
(b) $\beta_0 = 6.0, \xi_1 = 0.8$



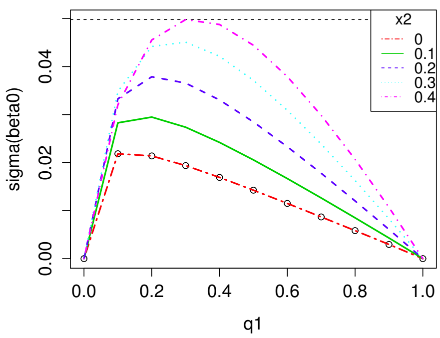
(c) $\beta_0 = 1.0, \xi_1 = 0.9$



(d) $\beta_0 = 6.0, \xi_1 = 0.9$

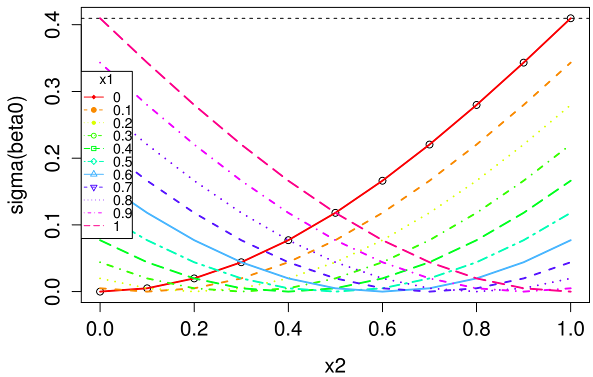


(e) $\beta_0 = 1.0, \xi_1 = 1.0$

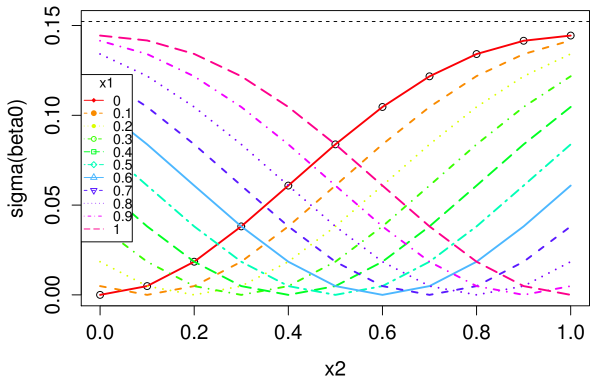


(f) $\beta_0 = 6.0, \xi_1 = 1.0$

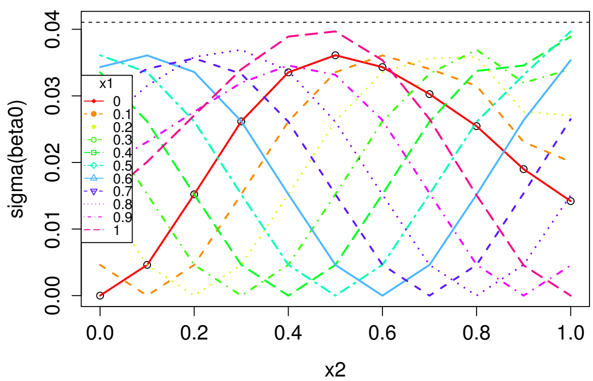
Figure A.3: The inverse of the asymptotic variance with $\beta_0 = 1.0$ (the left hand side) and $\beta_0 = 6.0$ (the right hand side) for the exponential model, $\lambda_0 = 1$, in pair (ξ_1, ξ_2) where $\xi_1 = 0.8, 0.9, 1$ and $\xi_2 = 0, 0.1, 0.2, 0.3, 0.4$ with the different q_1.



(a) $\beta_0 = 1.0$

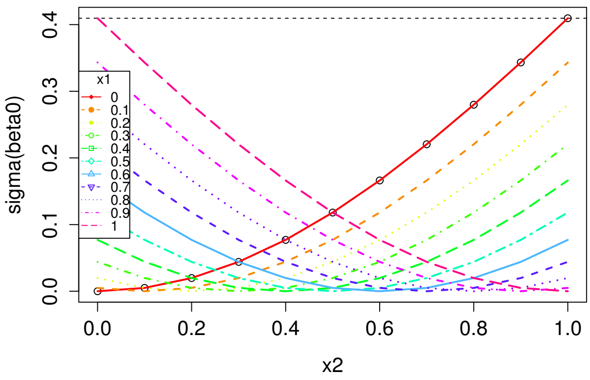


(b) $\beta_0 = 3.0$

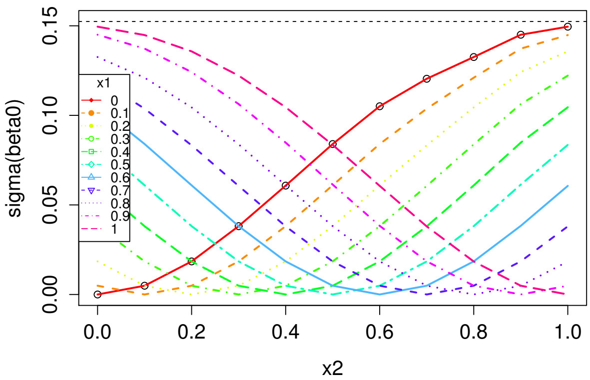


(c) $\beta_0 = 6.0$

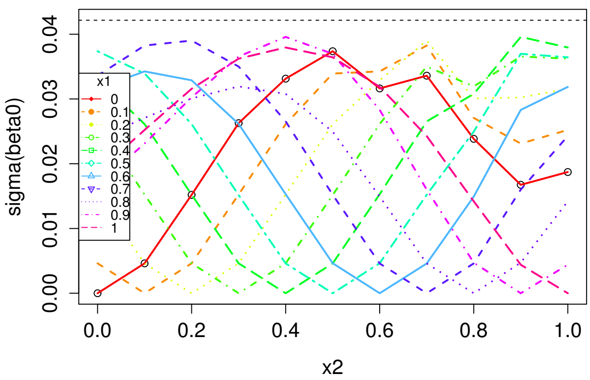
Figure A.4: The inverse of the asymptotic variance with β_0 for the Weibull model, $\theta_0 = 1, \mu = 2$, $q_1 = 0.5$ in every ξ_1 with the different ξ_2



(a) $\beta_0 = 1.0$

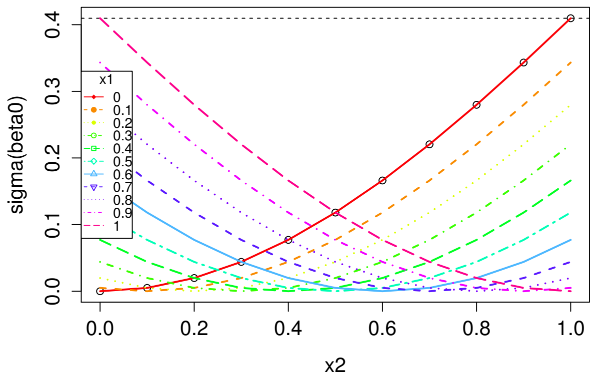


(b) $\beta_0 = 3.0$

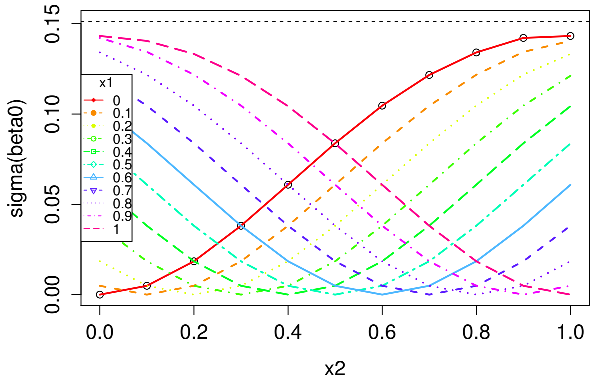


(c) $\beta_0 = 6.0$

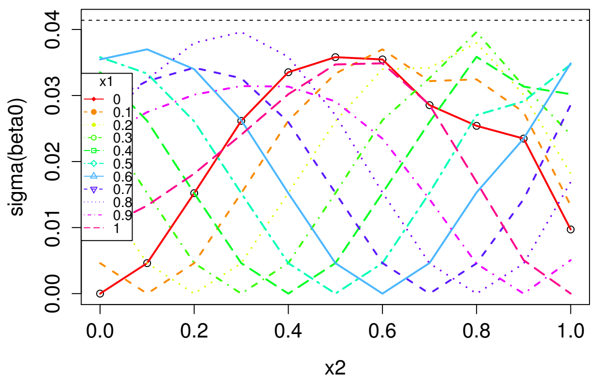
Figure A.5: The inverse of the asymptotic variance with β_0 for the Weibull model, $\theta_0 = 1, \mu = 5$, $q_1 = 0.5$ in every ξ_1 with the different ξ_2



(a) $\beta_0 = 1.0$

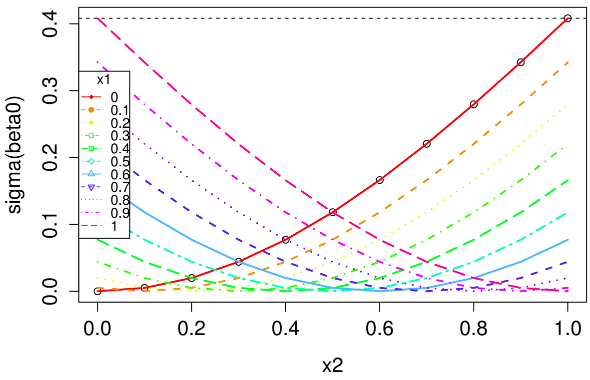


(b) $\beta_0 = 3.0$

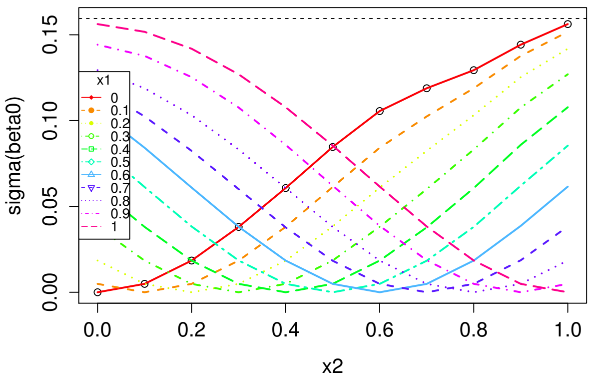


(c) $\beta_0 = 6.0$

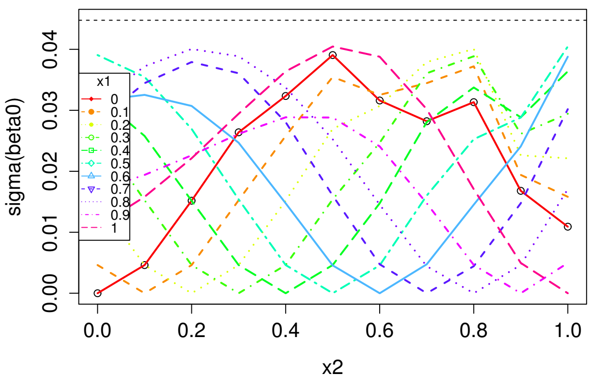
Figure A.6: The inverse of the asymptotic variance with β_0 for the Weibull model, $\theta_0 = 3, \mu = 2$, $q_1 = 0.5$, in every ξ_1 with the different ξ_2.



(a) $\beta_0 = 1.0$



(b) $\beta_0 = 3.0$



(c) $\beta_0 = 6.0$

Figure A.7: The inverse of the asymptotic variance with β_0 for the Weibull model, $\theta_0 = 3, \mu = 5$, $q_1 = 0.5$, in every ξ_1 with the different ξ_2.

Table A.1: Local optimal covariates for $\tilde{\beta}_0 = 1$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \delta_{[0,\tau)}(t)$ for $t \geq 0$.

(a,b,ω)	ξ_1^*	ξ_2^*	q_1^*	q_2^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\text{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	1	0.52431	0.47569	0.2052671	0.2052671	1
(1,0.3,1)	0	1	0.52431	0.47569	0.2052664	0.2052664	1
(1,0.6,1)	0	1	0.52431	0.47569	0.2052649	0.2052649	1
(1,0.9,1)	0	1	0.52431	0.47569	0.2052615	0.2052615	1
(1,0.75,15)	0	1	0.52431	0.47569	0.2048897	0.2048897	1
(1,0.9,30)	0	1	0.52431	0.47569	0.2081576	0.2081576	1
(2,0,1)	0	1	0.52431	0.47569	0.2052677	0.2052677	1
(2,0.3,1)	0	1	0.52431	0.47569	0.2052677	0.2052677	1
(2,0.6,1)	0	1	0.52431	0.47569	0.2052677	0.2052677	1
(2,0.9,1)	0	1	0.52431	0.47569	0.2052677	0.2052677	1
(2,0.75,15)	0	1	0.52431	0.47569	0.2052430	0.2052430	1
(2,0.9,30)	0	1	0.52431	0.47569	0.2052620	0.2052620	1
(5,0,1)	0	1	0.52431	0.47569	0.2052689	0.2052689	1
(5,0.3,1)	0	1	0.52431	0.47569	0.2052682	0.2052682	1
(5,0.6,1)	0	1	0.52431	0.47569	0.2052677	0.2052677	1
(5,0.9,1)	0	1	0.52431	0.47569	0.2052677	0.2052677	1
(5,0.75,15)	0	1	0.52431	0.47569	0.2052641	0.2052641	1
(5,0.9,30)	0	1	0.52431	0.47569	0.2056058	0.2056058	1
(10,0,1)	0	1	0.52431	0.47569	0.2052689	0.2052689	1
(10,0.3,1)	0	1	0.52431	0.47569	0.2052687	0.2052687	1
(10,0.6,1)	0	1	0.52431	0.47569	0.2052682	0.2052682	1
(10,0.9,1)	0	1	0.52431	0.47569	0.2052675	0.2052675	1
(10,0.75,15)	0	1	0.52431	0.47569	0.2052650	0.2052650	1
(10,0.9,30)	0	1	0.52431	0.47569	0.2052765	0.2052765	1

Table A.2: Local optimal covariates for $\tilde{\beta}_0 = 1$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \exp(-t)$.

(a,b,ω)	ξ_1^*	ξ_2^*	q_1^*	q_2^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\text{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	1	0.55290	0.44710	0.1372814	0.1372814	1
(1,0.3,1)	0	1	0.54850	0.45150	0.1481580	0.1481472	0.9999268
(1,0.6,1)	0	1	0.54500	0.45500	0.1567798	0.1562432	0.9997655
(1,0.9,1)	0	1	0.54225	0.45775	0.1624898	0.1624205	0.9995766
(1,0.75,15)	0	1	0.55345	0.44655	0.1382116	0.1382115	0.9999999
(1,0.9,30)	0	1	0.52431	0.47569	0.1375145	0.1375144	0.9999993
(2,0,1)	0	1	0.54080	0.45920	0.1648448	0.1647552	0.9994568
(2,0.3,1)	0	1	0.53905	0.46095	0.1691325	0.1690123	0.9992890
(2,0.6,1)	0	1	0.53760	0.46240	0.1726286	0.1724791	0.9991340
(2,0.9,1)	0	1	0.53640	0.46360	0.1755247	0.1753481	0.9989939
(2,0.75,15)	0	1	0.54125	0.45875	0.1659352	0.1658515	0.9994959
(2,0.9,30)	0	1	0.52431	0.47569	0.1656600	0.1655924	0.9995919
(5,0,1)	0	1	0.53155	0.46845	0.1870518	0.1867395	0.9983304
(5,0.3,1)	0	1	0.53120	0.46880	0.1879953	0.1876707	0.9982734
(5,0.6,1)	0	1	0.53085	0.46915	0.1888470	0.1885108	0.9982195
(5,0.9,1)	0	1	0.53055	0.46945	0.1896194	0.1892723	0.9981695
(5,0.75,15)	0	1	0.53165	0.46835	0.1879462	0.1876353	0.9983455
(5,0.9,30)	0	1	0.53180	0.46820	0.1876720	0.1873641	0.9983650
(10,0,1)	0	1	0.52805	0.47195	0.1957616	0.1953210	0.9977493
(10,0.3,1)	0	1	0.52795	0.47205	0.1960260	0.1955810	0.9977299
(10,0.6,1)	0	1	0.52785	0.47215	0.1962759	0.1958268	0.9977119
(10,0.9,1)	0	1	0.52775	0.47225	0.1965124	0.1960594	0.9976948
(10,0.75,15)	0	1	0.52795	0.47205	0.1962247	0.1957796	0.9977314
(10,0.9,30)	0	1	0.52810	0.47190	0.1960443	0.1956044	0.9977559

Table A.3: Local optimal covariates for $\tilde{\beta}_0 = 1$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = (\exp(ct) - 1)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}\log 2$.

(a,b,ω)	ξ_1^*	ξ_2^*	q_1^*	q_2^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\text{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	1	0.53075	0.46925	0.1906158	0.1906158	1
(1,0.3,1)	0	1	0.52940	0.47060	0.1937973	0.1937961	0.9999935
(1,0.6,1)	0	1	0.52845	0.47155	0.1959693	0.1959655	0.9999806
(1,0.9,1)	0	1	0.52770	0.47230	0.1974966	0.1974900	0.9999668
(1,0.75,15)	0	1	0.53075	0.46925	0.1904166	0.1904166	1
(1,0.9,30)	0	1	0.52700	0.47300	0.1924993	0.1924897	0.9999501
(2,0,1)	0	1	0.52725	0.47275	0.1981654	0.1981566	0.9999556
(2,0.3,1)	0	1	0.52685	0.47315	0.1990860	0.1990751	0.9999455
(2,0.6,1)	0	1	0.52655	0.47345	0.1998051	0.1997924	0.9999364
(2,0.9,1)	0	1	0.52630	0.47370	0.2003791	0.2003649	0.9999289
(2,0.75,15)	0	1	0.52735	0.47265	0.1982990	0.1982906	0.9999579
(2,0.9,30)	0	1	0.52940	0.47060	0.1983117	0.1983104	0.9999934
(5,0,1)	0	1	0.52540	0.47460	0.2024793	0.2024584	0.9998970
(5,0.3,1)	0	1	0.52535	0.47465	0.2026372	0.2026158	0.9998944
(5,0.6,1)	0	1	0.52530	0.47470	0.2027787	0.2027570	0.9998930
(5,0.9,1)	0	1	0.52525	0.47475	0.2029061	0.2028840	0.9998911
(5,0.75,15)	0	1	0.52545	0.47455	0.2026084	0.2025879	0.9998988
(5,0.9,30)	0	1	0.52555	0.47445	0.2028546	0.2028347	0.9999021
(10,0,1)	0	1	0.51535	0.48465	0.2039253	0.2038568	0.9996643
(10,0.3,1)	0	1	0.52480	0.47520	0.2039228	0.2038970	0.9998735
(10,0.6,1)	0	1	0.52480	0.47520	0.2039606	0.2039346	0.9998728
(10,0.9,1)	0	1	0.52480	0.47520	0.2039960	0.2039699	0.9998721
(10,0.75,15)	0	1	0.52485	0.47515	0.2039483	0.2039225	0.9998735
(10,0.9,30)	0	1	0.52485	0.47515	0.2039294	0.2039037	0.9998737

Table A.4: Local optimal covariates for $\tilde{\beta}_0 = 1$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = (ct)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}$.

(a,b,ω)	ξ_1^*	ξ_2^*	q_1^*	q_2^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\text{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	1	0.53280	0.46720	0.1854358	0.1854358	1
(1,0.3,1)	0	1	0.53115	0.46885	0.1895744	0.1895726	0.9999902
(1,0.6,1)	0	1	0.52990	0.47010	0.1924365	0.1924306	0.9999693
(1,0.9,1)	0	1	0.52895	0.47105	0.1944748	0.1944644	0.9999465
(1,0.75,15)	0	1	0.53290	0.46710	0.1853178	0.1853178	0.9999997
(1,0.9,30)	0	1	0.52950	0.47050	0.1870388	0.1870315	0.9999607
(2,0,1)	0	1	0.52835	0.47165	0.1953390	0.1953251	0.9999288
(2,0.3,1)	0	1	0.52785	0.47215	0.1965930	0.1965756	0.9999117
(2,0.6,1)	0	1	0.52745	0.47255	0.1975783	0.1975578	0.9998962
(2,0.9,1)	0	1	0.52710	0.47290	0.1983689	0.1983458	0.9998833
(2,0.75,15)	0	1	0.52850	0.47150	0.1955420	0.1955289	0.9999328
(2,0.9,30)	0	1	0.53050	0.46950	0.1955446	0.1955408	0.9999806
(5,0,1)	0	1	0.52590	0.47410	0.2012943	0.2012595	0.9998271
(5,0.3,1)	0	1	0.52580	0.47420	0.2015171	0.2014812	0.9998221
(5,0.6,1)	0	1	0.52570	0.47430	0.2017168	0.2016803	0.9998191
(5,0.9,1)	0	1	0.52565	0.47435	0.2018967	0.2018594	0.9998153
(5,0.75,15)	0	1	0.52595	0.47405	0.2014811	0.2014468	0.9998295
(5,0.9,30)	0	1	0.52600	0.47400	0.2016868	0.2016533	0.9998339
(10,0,1)	0	1	0.52510	0.47490	0.2032811	0.2032374	0.9997850
(10,0.3,1)	0	1	0.52505	0.47495	0.2033387	0.2032947	0.9997836
(10,0.6,1)	0	1	0.52505	0.47495	0.2033928	0.2033485	0.9997822
(10,0.9,1)	0	1	0.52500	0.47500	0.2034436	0.2033990	0.9997808
(10,0.75,15)	0	1	0.52505	0.47495	0.2033769	0.2033330	0.9997839
(10,0.9,30)	0	1	0.52505	0.47495	0.2033453	0.2033015	0.9997844

Table A.5: Local optimal covariates for $\tilde{\beta}_0 = 2$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \delta_{[0,\tau)}(t)$ for $t \geq 0$.

(a,b,ω)	ξ_1^*	ξ_2^*	q_1^*	q_2^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\text{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	1	0.52431	0.47569	0.1325058	0.1325058	1
(1,0.3,1)	0	1	0.52431	0.47569	0.1325034	0.1325034	1
(1,0.6,1)	0	1	0.52431	0.47569	0.1325054	0.1325054	1
(1,0.9,1)	0	1	0.52431	0.47569	0.1325056	0.1325056	1
(1,0.75,15)	0	1	0.52431	0.47569	0.1325819	0.1325819	1
(1,0.9,30)	0	1	0.52431	0.47569	0.1352414	0.1352414	1
(2,0,1)	0	1	0.52431	0.47569	0.1325058	0.1325058	1
(2,0.3,1)	0	1	0.52431	0.47569	0.1325059	0.1325059	1
(2,0.6,1)	0	1	0.52431	0.47569	0.1325034	0.1325034	1
(2,0.9,1)	0	1	0.52431	0.47569	0.1325008	0.1325008	1
(2,0.75,15)	0	1	0.52431	0.47569	0.1324563	0.1324563	1
(2,0.9,30)	0	1	0.52431	0.47569	0.1339720	0.1339720	1
(5,0,1)	0	1	0.52431	0.47569	0.1324532	0.1324532	1
(5,0.3,1)	0	1	0.52431	0.47569	0.1323959	0.1323959	1
(5,0.6,1)	0	1	0.52431	0.47569	0.1334675	0.1334675	1
(5,0.9,1)	0	1	0.52431	0.47569	0.1323023	0.1323023	1
(5,0.75,15)	0	1	0.52431	0.47569	0.1323630	0.1323630	1
(5,0.9,30)	0	1	0.52431	0.47569	0.1324197	0.1324197	1
(10,0,1)	0	1	0.52431	0.47569	0.1337034	0.1337034	1
(10,0.3,1)	0	1	0.52431	0.47569	0.1337085	0.1337085	1
(10,0.6,1)	0	1	0.52431	0.47569	0.1336715	0.1336715	1
(10,0.9,1)	0	1	0.52431	0.47569	0.1335915	0.1335915	1
(10,0.75,15)	0	1	0.52431	0.47569	0.1335659	0.1335659	1
(10,0.9,30)	0	1	0.52431	0.47569	0.1332663	0.1332663	1

Table A.6: Local optimal covariates for $\tilde{\beta}_0 = 3$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \delta_{[0,\tau)}(t)$ for $t \geq 0$.

(a,b,ω)	ξ_1^*	ξ_2^*	ξ_3^*	q_1^*	q_2^*	q_3^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\text{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	0.45	1	0.4178751	0.203	0.3791249	0.0795086	0.0734898	0.924299965
(1,0.3,1)	0	0.4	1	0.4183994	0.202	0.3796006	0.0794548	0.0734520	0.924450709
(1,0.6,1)	0	0.448	1	0.4178751	0.203	0.3791249	0.0795382	0.0735482	0.924690227
(1,0.9,1)	0	0.447	1	0.4299342	0.180	0.3900658	0.0794255	0.0736130	0.926818797
(1,0.75,15)	0	0.447	1	0.4173508	0.204	0.3786492	0.0794601	0.0734349	0.924172635
(1,0.9,30)	0	0.403	1	0.4220696	0.195	0.3829304	0.0837052	0.0782324	0.934618121
(2,0,1)	0	0.47	1	0.4199723	0.199	0.3810277	0.0792923	0.0735694	0.927825229
(2,0.3,1)	0	0.44	1	0.4241668	0.191	0.3848332	0.0794301	0.0732608	0.922330452
(2,0.6,1)	0	0.461	1	0.4163021	0.206	0.3776979	0.0793825	0.0729614	0.919111895
(2,0.9,1)	0	0.458	1	0.4168265	0.205	0.3781736	0.0795624	0.0729680	0.917116
(2,0.75,15)	0	0.456	1	0.4168265	0.205	0.3781736	0.0795447	0.0731696	0.919855075
(2,0.9,30)	0	0.463	1	0.4189237	0.201	0.3800763	0.0797236	0.0735094	0.922053772
(5,0,1)	0	0.423	1	0.4267883	0.186	0.3872117	0.0792677	0.0752376	0.949158326
(5,0.3,1)	0	0.439	1	0.4252154	0.189	0.3857846	0.0787294	0.0796202	0.947805996
(5,0.6,1)	0	0.463	1	0.4225939	0.194	0.3834061	0.0781631	0.0787703	0.943799568
(5,0.9,1)	0	0.49	1	0.4204966	0.198	0.3815034	0.0775997	0.0727626	0.937665996
(5,0.75,15)	0	0.475	1	0.4210209	0.197	0.3819791	0.0778905	0.0732300	0.94016536
(5,0.9,30)	0	0.482	1	0.4183994	0.202	0.3796006	0.0777377	0.0776719	0.934191476
(10,0,1)	0	0.482	1	0.3942811	0.248	0.3577189	0.0802415	0.0682332	0.850347919
(10,0.3,1)	0	0.473	1	0.3942811	0.248	0.3577189	0.0809344	0.0689444	0.85185533
(10,0.6,1)	0	0.465	1	0.3948054	0.247	0.3581946	0.0816313	0.0697399	0.854327846
(10,0.9,1)	0	0.456	1	0.3953297	0.246	0.3586703	0.0823594	0.0706053	0.857282229
(10,0.75,15)	0	0.460	1	0.3948054	0.247	0.3581946	0.0820100	0.0701889	0.855857213
(10,0.9,30)	0	0.456	1	0.3953297	0.246	0.3586703	0.0823888	0.0707084	0.858228217

Table A.7: Local optimal covariates for $\tilde{\beta}_0 = 3$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = 1 - \exp(-t)$.

(a,b,ω)	ξ_1^*	ξ_2^*	ξ_3^*	q_1^*	q_2^*	q_3^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\mathrm{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	0.545	1	0.5964651	0.094	0.3095349	0.0717005	0.0705166	0.983488
(1,0.3,1)	0	0.542	1	0.5855980	0.108	0.3064020	0.0734779	0.0718965	0.978478
(1,0.6,1)	0	0.538	1	0.5777824	0.116	0.3062176	0.0746761	0.0728204	0.975151
(1,0.9,1)	0	0.564	1	0.5731880	0.120	0.3068120	0.0754256	0.0733336	0.972265
(1,0.75,15)	0	0.535	1	0.5999136	0.088	0.3120864	0.0731693	0.0721655	0.986280
(1,0.9,30)	0	0.498	1	0.6196014	0.059	0.3213986	0.0759979	0.0755740	0.994423
(2,0,1)	0	0.545	1	0.5433093	0.166	0.2906907	0.0754934	0.0734375	0.972768
(2,0.3,1)	0	0.553	1	0.5792150	0.129	0.2917850	0.0762074	0.0736908	0.966977
(2,0.6,1)	0	0.519	1	0.5692997	0.135	0.2957002	0.0768280	0.0740581	0.963947
(2,0.9,1)	0	0.520	1	0.5710837	0.134	0.2949163	0.0773388	0.0746159	0.964793
(2,0.75,15)	0	0.538	1	0.5767152	0.123	0.3002848	0.0766177	0.0743539	0.970453
(2,0.9,30)	0	0.683	1	0.5882798	0.102	0.3097202	0.0757854	0.0742788	0.980120
(5,0,1)	0	0.532	1	0.5347417	0.138	0.3272583	0.0780655	0.0753487	0.965198
(5,0.3,1)	0	0.557	1	0.5293076	0.144	0.3266924	0.0777050	0.0746038	0.960133
(5,0.6,1)	0	0.581	1	0.5270668	0.149	0.3239332	0.0773470	0.0737927	0.954048
(5,0.9,1)	0	0.588	1	0.5163796	0.157	0.3266204	0.0770111	0.0729853	0.947724
(5,0.75,15)	0	0.584	1	0.5210744	0.153	0.3259256	0.0771411	0.0733879	0.951346
(5,0.9,30)	0	0.590	1	0.5296522	0.156	0.3143478	0.0771083	0.0730618	0.947522
(10,0,1)	0	0.506	1	0.5944365	0.165	0.2405635	0.0810039	0.0753487	0.927722
(10,0.3,1)	0	0.496	1	0.5975760	0.160	0.2424240	0.0815929	0.0761100	0.932801
(10,0.6,1)	0	0.487	1	0.6003302	0.155	0.2446697	0.0822088	0.0771105	0.937983
(10,0.9,1)	0	0.477	1	0.6034866	0.149	0.2475133	0.0828482	0.0781392	0.943162
(10,0.75,15)	0	0.482	1	0.6018680	0.152	0.2461320	0.0825270	0.0776339	0.940709
(10,0.9,30)	0	0.478	1	0.6031037	0.149	0.2478963	0.0828508	0.0781870	0.943709

Table A.8: Local optimal covariates for $\tilde{\beta}_0 = 3$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = (\exp(ct) - 1)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}\log 2$.

(a,b,ω)	ξ_1^*	ξ_2^*	ξ_3^*	q_1^*	q_2^*	q_3^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\mathrm{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	0.531	1	0.5537676	0.147	0.2992324	0.0791761	0.0758667	0.958202
(1,0.3,1)	0	0.527	1	0.5520325	0.150	0.2979675	0.0795445	0.0760935	0.956615
(1,0.6,1)	0	0.529	1	0.5500247	0.151	0.2989754	0.0797670	0.0762904	0.956416
(1,0.9,1)	0	0.517	1	0.5460390	0.155	0.2989610	0.0799154	0.0762917	0.954656
(1,0.75,15)	0	0.526	1	0.5538102	0.147	0.2991897	0.0793458	0.0760875	0.958936
(1,0.9,30)	0	0.494	1	0.5751173	0.115	0.3098828	0.0830659	0.0811518	0.976957
(2,0,1)	0	0.520	1	0.5455616	0.156	0.2984384	0.0802953	0.0762491	0.949609
(2,0.3,1)	0	0.561	1	0.5801950	0.153	0.2668050	0.0795960	0.0761200	0.956329
(2,0.6,1)	0	0.516	1	0.5517435	0.157	0.2912565	0.0801589	0.0761822	0.950390
(2,0.9,1)	0	0.517	1	0.5554413	0.154	0.2905587	0.0803399	0.0765158	0.952401
(2,0.75,15)	0	0.526	1	0.5532840	0.154	0.2927160	0.0801600	0.0764552	0.953782
(2,0.9,30)	0	0.534	1	0.5510158	0.153	0.2959842	0.0803716	0.0766009	0.953085
(5,0,1)	0	0.522	1	0.5247245	0.151	0.3242756	0.0798937	0.0767113	0.960167
(5,0.3,1)	0	0.545	1	0.5197352	0.156	0.3242648	0.0794380	0.0758870	0.955298
(5,0.6,1)	0	0.570	1	0.5179440	0.160	0.3220560	0.0789839	0.0749855	0.949377
(5,0.9,1)	0	0.580	1	0.5085882	0.167	0.3244119	0.0785614	0.0740799	0.942955
(5,0.75,15)	0	0.573	1	0.5125934	0.164	0.3234066	0.0787765	0.0745313	0.946111
(5,0.9,30)	0	0.569	1	0.5093295	0.170	0.3206705	0.0787150	0.0741557	0.942121
(10,0,1)	0	0.506	1	0.5891220	0.172	0.2388780	0.0869054	0.0752968	0.919314
(10,0.3,1)	0	0.497	1	0.5930574	0.166	0.2409426	0.0824615	0.0762405	0.924563
(10,0.6,1)	0	0.487	1	0.5958578	0.161	0.2431422	0.0830485	0.0772295	0.929932
(10,0.9,1)	0	0.478	1	0.5990205	0.155	0.2459795	0.0836597	0.0782512	0.935351
(10,0.75,15)	0	0.482	1	0.5973990	0.158	0.2446010	0.0833573	0.0777503	0.932735
(10,0.9,30)	0	0.478	1	0.5986402	0.155	0.2463597	0.0836794	0.0783055	0.935779

Table A.9: Local optimal covariates for $\tilde{\beta}_0 = 3$ and efficiencies for different $\tilde{\lambda}_0$ with censoring distribution $G(t) = (ct)\delta_{[0,\tau)}(t) + \delta_{[\tau,\infty)}(t)$ for $c = \frac{1}{\tau}$.

(a,b,ω)	ξ_1^*	ξ_2^*	ξ_3^*	q_1^*	q_2^*	q_3^*	$\mathbf{\Sigma}^*(\tilde{\beta}_0, \tilde{\lambda}_0, G)$	$\mathbf{\Sigma}(\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0, G, Q^*(\boldsymbol{\beta}_0, \lambda_0))$	$\text{eff}_{\tilde{\boldsymbol{\beta}}_0, \tilde{\lambda}_0}(Q^*(\boldsymbol{\beta}_0, \lambda_0))$
(1,0,1)	0	0.532	1	0.5569643	0.143	0.3000357	0.0765489	0.0754605	0.960682
(1,0.3,1)	0	0.529	1	0.5544073	0.147	0.2985926	0.0790483	0.0757796	0.958649
(1,0.6,1)	0	0.529	1	0.5516608	0.149	0.2993393	0.0793630	0.0760332	0.958044
(1,0.9,1)	0	0.518	1	0.5482744	0.152	0.2997256	0.0795682	0.0760738	0.956083
(1,0.75,15)	0	0.527	1	0.5569643	0.143	0.3000357	0.0788102	0.0757899	0.961676
(1,0.9,30)	0	0.495	1	0.5789005	0.110	0.3110995	0.0824698	0.0764303	0.926767
(2,0,1)	0	0.521	1	0.5478396	0.153	0.2991604	0.0799733	0.0760423	0.950847
(2,0.3,1)	0	0.560	1	0.5804188	0.151	0.2685811	0.0793411	0.0759421	0.957159
(2,0.6,1)	0	0.517	1	0.5526090	0.156	0.2913910	0.0799028	0.0760276	0.951500
(2,0.9,1)	0	0.517	1	0.5562672	0.153	0.2907327	0.0801120	0.0763783	0.953393
(2,0.75,15)	0	0.527	1	0.5555007	0.151	0.2934993	0.0798799	0.0763021	0.955210
(2,0.9,30)	0	0.535	1	0.5532225	0.150	0.2967775	0.0800695	0.0764303	0.954550
(5,0,1)	0	0.522	1	0.5254700	0.150	0.3245300	0.0797600	0.0766129	0.960543
(5,0.3,1)	0	0.546	1	0.5205200	0.155	0.3244800	0.0793089	0.0757941	0.955682
(5,0.6,1)	0	0.571	1	0.5187288	0.159	0.3222712	0.0788618	0.0748994	0.949755
(5,0.9,1)	0	0.581	1	0.5093238	0.166	0.3246762	0.0784442	0.0740009	0.943357
(5,0.75,15)	0	0.574	1	0.5133321	0.163	0.3236679	0.0786542	0.0744487	0.946532
(5,0.9,30)	0	0.570	1	0.5100678	0.169	0.3209322	0.0785855	0.0740768	0.942627
(10,0,1)	0	0.506	1	0.5898750	0.171	0.2391250	0.0818393	0.0752890	0.919961
(10,0.3,1)	0	0.497	1	0.5930574	0.166	0.2409426	0.0823986	0.0762342	0.925188
(10,0.6,1)	0	0.487	1	0.5965680	0.160	0.2434320	0.0829878	0.0772239	0.930546
(10,0.9,1)	0	0.478	1	0.5990205	0.155	0.2459795	0.0836011	0.0782463	0.935948
(10,0.75,15)	0	0.482	1	0.5973990	0.158	0.2446010	0.0832973	0.0777450	0.933344
(10,0.9,30)	0	0.478	1	0.5986402	0.155	0.2463597	0.0836195	0.0783000	0.936384

Bibliography

P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, 10(4):1100–1120, 1982.

P. K. Andersen, Borgan, R. D. Gill, and N. Keiding. *Statistical Models based on Counting Processes*. Springer-Verlag New York, 1993.

J. Arnlöv, E. Ingelsson, J. Sundström, and L. Lind. Impact of body mass index and the metabolic syndrome on the risk of cardiovascular disease and death in middle-aged men. *Circulation*, 121:230–236, 2010.

N. Balakrishnan and Donghoon Han. Optimal progressive type-II censoring schemes for nonparametric confidence intervals of quantiles. *Comm. Statist. Simulation Comput.*, 36(4-6):1247–1262, 2007.

N. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30 (1):89–99, 1974.

Z. Cai and Y. Sun. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scand. J. Statist.*, 30(1):93–111, 2003.

J. Cologne, DL. Preston, K. Imai, M. Misumi, K. Yoshida, T. Hayashi, and K. Nakachi. Conventional case-cohort design and analysis for studies of interaction. *Int J Epidemiol*, 41(4):1174–1186, 2012.

D. R. Cox. Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34:187–220, 1972.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge, 1997.

B. Efron. The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565, 1977.

J. Fan, I. Gijbels, and M. King. Local likelihood and local partial likelihood in hazard regression. *Ann. Statist.*, 25(4):1661–1690, 1997.

V. V. Fedorov. *Theory of optimal experiments.* Academic Press, New York-London, 1972. Translated from the Russian and edited by W. J. Studden and E. M. Klimko, Probability and Mathematical Statistics, No. 12.

S. Garcet-Rodríguez, J. López-Fidalgo, and R. Martín-Martín. Some complexities in optimal experimental designs introduced by real life problems. *Tatra Mt. Math. Publ.*, 39:135–143, 2008.

T. Hastie and R. Tibshirani. Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B*, 55(4):757–796, 1993. With discussion and a reply by the authors.

J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data.* Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2002.

G. Kauermann and U. Berger. A smooth test in proportional hazard survival models using local partial likelihood fitting. *Lifetime Data Anal.*, 9(4):373–393, 2003.

J. Kiefer. Optimum designs in regression problems. II. *Ann. Math. Statist.*, 32:298–325, 1961.

D. G. Kleinbaum and M. Klein. *Survival Analysis: A Self-Learning Text.* Springer Science+Business Media, Inc., New York, 2005.

E. Läuter and H. Läuter. Experimental design methods. *Statistical inference in linear models (ed. by H. Bunke, O. Bunke),John Wiley, Chichester*, pages 505–568, 1984.

J. López-Fidalgo, M. J. Rivas-López, and R. del Campo. Optimal designs for Cox regression. *Stat. Neerl.*, 63(2):135–148, 2009.

T. Martinussen and T. H. Scheike. A nonparametric dynamic additive regression model for longitudinal data. *Ann. Statist.*, 28(4):1000–1025, 2000.

T. Martinussen and T. H. Scheike. *Dynamic Regression Models for Survival Data.* Springer-Verlag New York, 1st edition, 2006.

L. Marzec and P. Marzec. On fitting cox's regression model with time-dependent coefficients. *Biometrika*, 84(4):901–908, 1997.

S. A. Murphy. Testing for a time dependent coefficient in Cox's regression model. *Scand. J. Statist.*, 20(1):35–50, 1993.

S. A. Murphy and P. K. Sen. Time-dependent coefficients in a Cox-type regression model. *Stochastic Process. Appl.*, 39(1):153–180, 1991.

R. L. Prentice. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73(1):1–11, 1986.

T. H. Scheike and T. Martinussen. On estimation and tests of time-varying effects in the proportional hazards model. *Scand. J. Statist.*, 31(1):51–62, 2004.

D. Schmidt and R. Schwabe. On optimal designs for censored data. *Metrika*, 78(3):237–257, 2015.

T. M. Therneau and P. M. Grambsch. *Modeling Survival Data: Extending the Cox Model.* Springer, New York, 2000. ISBN 0-387-98784-3.

L. Tian, D. Zucker, and L. J. Wei. On the Cox model with time-varying regression coefficients. *J. Amer. Statist. Assoc.*, 100(469):172–183, 2005. ISSN 0162-1459. doi: 10.1198/016214504000000845. URL `http://dx.doi.org/10.1198/016214504000000845`.

R. Tibshirani and T. Hastie. Local likelihood estimation. *Journal of the American Statistical Association*, 82(398):559–567, 1987.

P. Whittle. Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B*, 35:123–130, 1973.

H. P. Wynn. Results in the theory and construction of d-optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):pp. 133–147, 1972.

D. M. Zucker and A. F. Karr. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.*, 18(1):329–353, 1990.