Optimal Design for Dose Finding Studies on Safety and Efficacy

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von Dipl.-Stat. Katrin Roth
geb. am 16.02.1981 in Herdecke

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Rainer Schwabe
Prof. Dr. Frank Bretz

eingereicht am: 19.06.2009
Verteidigung am: 19.11.2009

Acknowledgements

I would like to use this opportunity to thank some of the people without whose help this PhD thesis would not have been possible.

First of all I want to thank my advisor Prof. Dr. Rainer Schwabe for his constant support, his patience with my questions, the helpful discussions and his valuable suggestions and comments.

Special thanks go to Dr. Thomas Schmelter. Sharing his experiences and knowledge, taking the time to listen to my problems, encouraging, motivating and pushing me, he was very helpful throughout the last years. I am very grateful for that.

Moreover I would like to thank Dr. Hermann Kulmann and the whole 'Clinical Statistics Europe I' department at Bayer Schering Pharma for the nice atmosphere and for immediately integrating me into all departmental activities. This gave me the opportunity to not only work on my PhD thesis surrounded by many helpful experienced statisticians, but also to learn a lot about the daily work in a pharmaceutical company.

My thanks also go to the 'Institut für Mathematische Stochastik' at the University of Magdeburg. Even though I was not around that much, the team provided me with all the support I needed and made the trips to Magdeburg more worthwhile.

I would like to thank all my friends for the refreshing distraction from work and the great time spent together.

Last but not least I want to thank my parents for always supporting me, trusting in me and letting me go my own way.

I am much obliged to Bayer Schering Pharma AG for the financial support of this project.

Summary

This thesis deals with optimal experimental design theory, applied to dose finding studies.

After a brief introduction to the field of clinical studies and dose finding studies in particular, an introduction to the well-known general optimal design theory, especially for nonlinear models, is given.

Subsequently, commonly used models for dose finding studies (the ordinary logistic model, the proportional odds model and the $E_{\max}$ model) are presented. For the logistic and the proportional odds model, conditions for the existence of the maximum likelihood estimator are derived and optimal designs are developed for these models.

After this, a new model for two response variables, one of them categorical and the other one binary, is introduced. It is applied to model toxicity and efficacy simultaneously, accounting for possible dependencies of both response variables. First the information matrix is derived for this model, which is then used to exemplarily determine locally D-optimal designs for specific parameter settings.

Furthermore, we suggest the 'sequential locally optimal design' ('SLOD') as a sequential approach combining a simple standard method for dose escalation studies (3+3 design) with the results of optimal design theory. For the variables that are of interest in this approach, variances and corresponding confidence intervals are derived.

This approach is compared to the standard method and a Bayesian approach with respect to its behavior in realistic study settings within a simulation study. The simulations show the advantages of the new suggested approach over the existing methods.

We conclude with a discussion of the results and an outlook to possible future work.

Zusammenfassung

Diese Arbeit behandelt die Theorie der optimalen Versuchsplanung, welche auf Dosisfindungsstudien angewendet wird.

Nach einer kurzen Einführung in das Gebiet der klinischen Studien, insbesondere der Dosisfindungsstudien, wird eine Einführung in die allgemein bekannte Theorie der optimalen Versuchsplanung gegeben. Dabei wird speziell auf den Fall von nichtlinearen Modellen eingegangen.

Nachfolgend werden häufig für Dosisfindungsstudien verwendete Modelle (das logistiche Modell, das Proportional-Odds-Modell und das $E_{\max}$-Modell) vorgestellt. Für das logistische und das Proportional-Odds-Modell werden Bedingungen für die Existenz des Maximum-Likelihood-Schätzers hergeleitet, und die Theorie der optimalen Versuchsplanung wird auf diese Modelle angewandt.

Danach wird ein neues Modell für zwei Zielvariablen, von denen eine kategoriell und die andere binär ist, eingeführt. Dieses wird zur simultanen Modellierung der Toxizität und der Wirksamkeit unter Berücksichtigung möglicher Abhängigkeiten zwischen diesen Zielvariablen verwendet. Zunächst wird die Informationsmatrix für dieses Modell hergeleitet, die dann zur Bestimmung beispielhafter lokal D-optimaler Versuchspläne für bestimmte Parameter verwendet wird.

Des Weiteren wird mit dem „Sequential Locally Optimal Design“, kurz „SLOD“, ein sequentieller Ansatz zur Kombination der einfachen Standardmethode für Dosis-Eskalations-Studien (dem 3+3-Design) und der optimalen Versuchsplanung eingeführt. Varianzen und zugehörige Konfidenzintervalle für die interessierenden Größen bei diesem Ansatz werden hergeleitet.

Dieser Ansatz wird mit Hilfe einer Simulationsstudie mit der Standardmethode und einem Bayesianischen Ansatz bezüglich des Verhaltens in realistischen Studiensituationen verglichen. Die Simulationsstudie zeigt die Vorteile des neu vorgeschlagenen Ansatzes gegenüber den bekannten Methoden.

Die Arbeit schließt mit einer Diskussion der Ergebnisse und einem Ausblick auf mögliche zukünftige Arbeiten.

Contents

List of Tables

List of Figures

1 Introduction

The theory of optimal experimental design is well-developed for ordinary linear models. The most important results on convex design theory date back about fifty years, comprising important equivalence theorems (e.g. Kiefer (1959) and Kiefer and Wolfowitz (1960)).

These results have been extended and generalized to nonlinear models (cf. White (1973)). However, the results on optimal design for nonlinear models, which we will use throughout this work, are less comprehensive. The same holds for sequential approaches, which will also be of importance within this work. Both topics are touched e.g. in Silvey (1980).

In this thesis, the theory of optimal design is applied to the field of clinical studies. Design of experiments plays an important role in this field because we are dealing with controlled experiments. Additionally, special issues arise due to the fact that we are dealing with experiments conducted in humans. We have to take into consideration ethical aspects like not exposing patients to toxic substances or treating patients with inefficacious drugs.

We will confine ourselves to the early phase clinical trials where the drug has not yet been tested in many people or where it is even the first time the drug is applied in humans. The goal of these studies is finding an appropriate dose of the drug that is both safe and efficacious. In these early phases, studies are often conducted sequentially, such that the outcomes of the previous patients determine which dose of the drug is administered to the successive patients. The standard methods in this field comprise simple up-and-down designs like the 3+3 design (cf. Ivanova (2006a) and Lin and Shih (2001)). These designs are easy to apply in practice, but are not optimal in the sense of design theory. They have been discussed and extended, e.g. by Ivanova (2006b).

Other more sophisticated approaches like the continual reassessment method (cf. O'Quigley et al. (1990)), purely Bayesian approaches (e.g. Whitehead and Williamson (1998)), designs based on bivariate models considering both toxicity and efficacy (cf. Dragalin et al. (2006) and Dragalin and Fedorov (2006)), on efficacy-toxicity trade-offs (cf. Thall and Cook (2004)) or on contingent response models (cf. Rabie and Flournoy (2004)) were suggested over the past years. They are mostly based on specific models and binary response variables, which is a major restriction.

The purpose of this work is to improve the designs for dose finding studies, in particular where toxicity and efficacy are considered simultaneously. We will introduce appropriate models suitable for categorical response variables and derive locally optimal designs for these models. Subsequently we will transfer these results to a more applicable sequential setting. Additionally we want to investigate the practical feasibility of these designs and compare them to standard methods.

In Chapter 2 we will give a brief introduction to the drug development process and dose finding studies. We will continue with an introduction to optimal design theory in Chapter 3. There we summarize some of the existing results that we will use in the subsequent chapters. We will present some commonly used models for dose finding studies in Chapter 4, namely the logistic, the proportional odds and the $E_{\max}$ model. Subsequently in Chapter 5 we will introduce a bivariate model applicable to dose finding studies with two endpoints, i.e., toxicity and efficacy, and we will derive locally optimal designs for this model. In Chapter 6 we introduce the 'sequential locally optimal design' ('SLOD'), a specific sequential approach for the conduct of dose escalation studies. We will carry out simulations to explore the properties of this new approach and to compare it to other methods in Chapter 7. We will conclude with a discussion of the results in Chapter 8.

2 Introduction to Dose Finding Studies

This chapter serves to describe the background of dose finding studies as part of the drug development process. The most important terms and definitions are presented and the general concepts are explained. The biological, medical and pharmaceutical details are of minor importance, the focus is on the statistical methods used in designing and analyzing dose finding studies. Additionally we will discuss the specific challenges in this field of application. First, a rough draft of the general drug development process is given. After that, the different types of dose finding studies are described, followed by a presentation of some methods specific to this field. The descriptions in this chapter are based on Ting (2006b), Modi (2006), Ivanova (2006a), Tighiouart and Rogatko (2006) and MacDougall (2006).

2.1 Drug Development Process

Before a new drug is made available to the public, its effects on the human organism have to be studied intensively. Not only the desired effects of the drug in curing the disease under consideration, but also the unwanted effects – so called adverse events – are of interest, since the goal is to develop a drug that is efficacious in curing the disease and at the same time safe in the sense that it does not have severe adverse effects.

The complete drug development process from a new chemical compound to a drug available on the market consists of two main parts, the nonclinical and the clinical development. The nonclinical development includes all drug development activities and experiments that are not performed in humans, e.g. in-vitro studies and studies in animals. The term clinical development or clinical studies refers to drug testing conducted within the human body. For the following we will focus on the clinical development, although some of the proposed methods might as well be applied to animal studies.

The clinical development consists of four different phases called Phase I, II, III and IV. The first three phases are performed before a drug is marketed, Phase IV studies are conducted when the drug is already available on the market.

Phase I studies aim at collecting information about the effects of the drug on the human body and vice versa. It is of major interest, how the drug is absorbed by the human body and how the body reacts to the drug. The focus is also on determining which doses are tolerated by the human body. Phase I studies are usually performed in healthy volunteers and are small-scale (i.e. comprising only few subjects) and short-term.

In Phase II trials, which are usually carried out in patients suffering from the target disease, the efficacy of the drug is to be assessed. These studies are still quite small-scale and serve to explore the dose response relationship in more detail, with respect to both the efficacy as well as the tolerability of the drug in a well defined patient population.

Phase III trials are usually meant to confirm the observations and results from the earlier studies. They compare the new drug to a reference drug or to placebo. They are of much larger scale and longer term than the earlier phase studies and the patient population is less restricted.

We are mainly interested in Phase I/II studies, with special focus on investigating the dose response relationship for both the safety and the efficacy of the drug.

2.2 Dose Finding Studies

There are different types of dose finding studies performed within Phase I and Phase II. The first studies related to dose finding are dose escalation studies. They aim at finding the maximally tolerated dose. Later on the goal is to determine the therapeutic window, that is, to find doses that are both efficacious and safe. Efficacy often is established in comparison to placebo.

2.2.1 Dose Escalation Studies

The so called dose escalation studies are Phase I studies and are among the first studies of the new compound carried out in humans. Throughout the course of such a study, the doses considered are escalated from a low starting dose to a higher target dose. They usually are performed with healthy volunteers, but may also be conducted on patients suffering from the target disease. The latter is especially done in oncology, when the disease is life-threatening and the drug is expected to cause major adverse events, which are partly tolerated due to the severeness of the disease. The goal of dose escalation studies is to collect information about the dose-toxicity-relationship, especially to determine a dose that is considered to be the maximum tolerated dose (MTD). The exact definition of the MTD depends on the disease and the treatment under consideration.

Dose escalation studies comprise only few subjects. The focus is on the safety of the new compound, the efficacy is not yet of major interest.

In oncology studies in patients suffering from a life threatening disease, the MTD is typically defined as the dose expected to cause unacceptable toxic events in a certain proportion of the patient population. The unacceptable toxic events are called dose limiting toxicities (DLTs) and are usually based on a toxicity grading using the Common Terminology Criteria for Adverse Events of the National Cancer Institute (cf National Cancer Institute (2006)). Define Γ as the maximally tolerated level of toxicity. The MTD then is the highest dose with probability of DLT less than or equal to Γ. Γ is often specified as a value around 0.33.

Before the conduct of the first dose escalation studies, there is very few or no prior knowledge available how humans will react to the new drug, because the information gathered in nonclinical studies is hard to transfer to humans. The challenge therefore is to get reliable results about the dose-toxicity-relationship and the MTD without treating patients at highly toxic doses and by using as few subjects as possible.

2.2.2 Dose Response Studies

With the term dose response studies we summarize all other dose finding studies except for the dose escalation studies described above. They are mainly performed in Phase II and aim at collecting information about the relationship between the dose and the efficacy response as well as the adverse effects. Their main purpose usually is estimating the dose response relationship, the optimal dose or the therapeutic window. The upper bound of the therapeutic window is given by the MTD established in Phase I. For the lower bound, a minimum effective dose ($minED$) can be defined. Often the minimum effective dose is defined as the minimum dose that induces a clinically relevant effect in a certain proportion of the patient population. The optimal dose can be defined as the dose that maximizes the joint probability of efficacy and no toxicity.

For illustration an example of a dose response relationship for efficacy and toxicity is shown in Figure 2.1. The therapeutic window is also displayed there. For this example, it is given by the doses for which the probability of efficacy is greater than 0.8 and the probability of toxicity is less than 0.33.

2.3 The 3+3 Design

The 3+3 design, also known as the traditional escalation rule, is a design widely used in Phase I dose escalation studies, especially in oncology. Before the start of the trial, a sequence of doses is specified. The starting dose is deduced from the

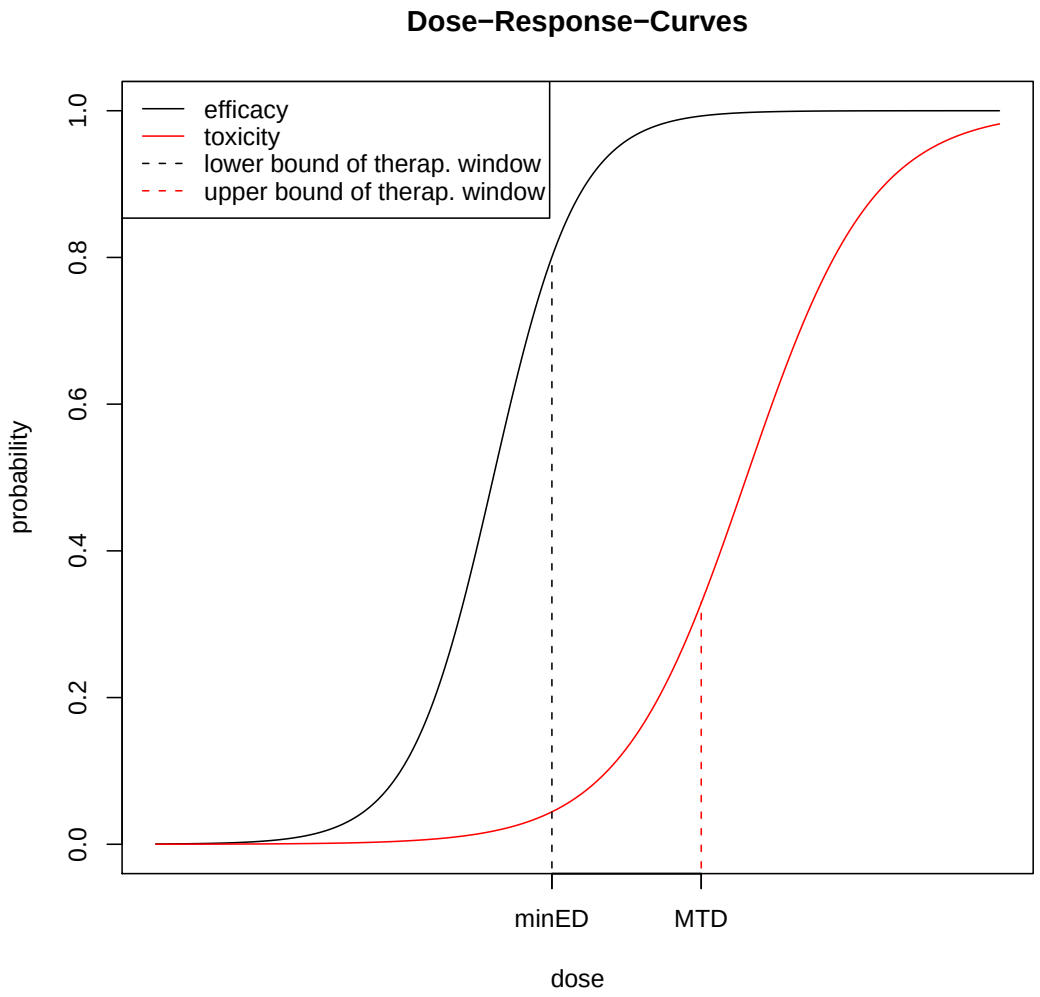


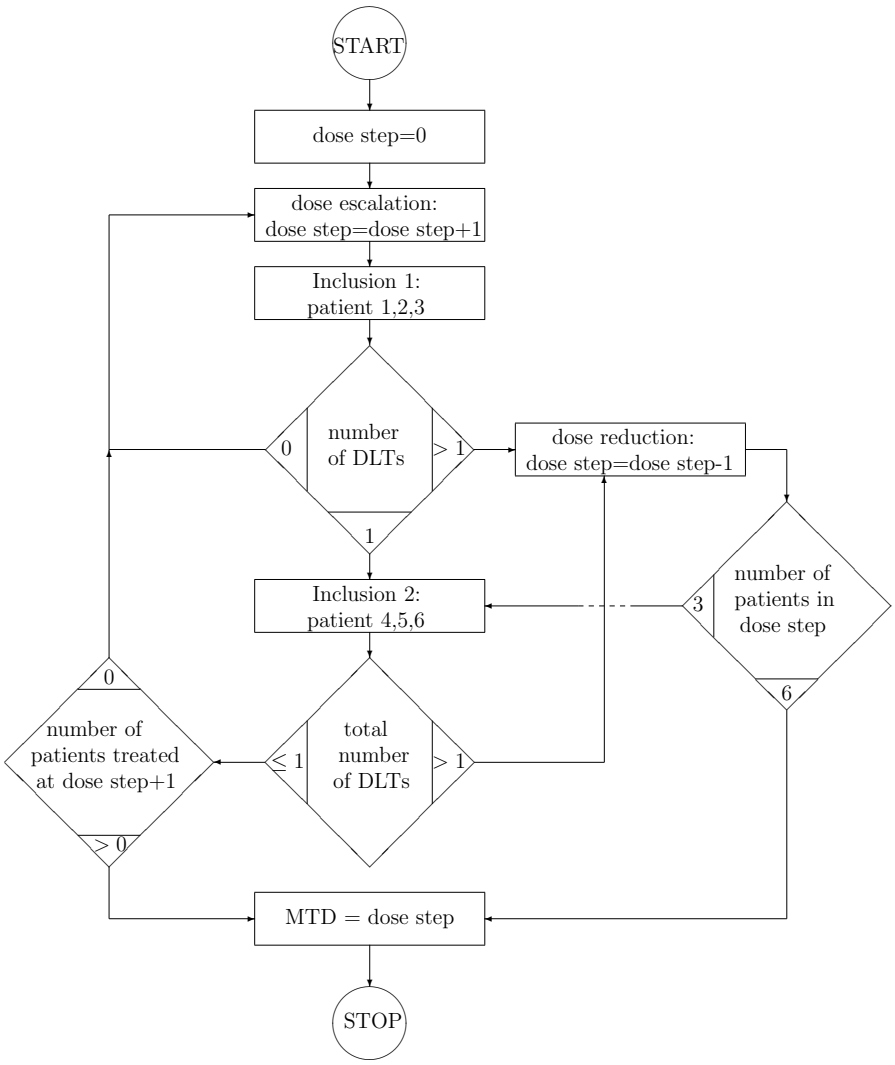
Figure 2.1: Example of a dose response relationship and the therapeutic window.

results of animal studies. The subsequent doses are determined by increasing the previous dose by 100%, then 65%, then 50, 40 and thereafter always 33%. The outcome considered is the occurrence of DLTs and therefore is binary. Subjects are treated in cohorts of three, receiving the same dose. The first three subjects are treated with the lowest dose. The treatment of the next cohort depends on the outcome of the previous cohorts. If three patients have been treated at a dose so far, the next cohort of patients is treated at the next higher dose level, if no toxicities are observed in this cohort. If one out of the three patients experiences a DLT, the next cohort is assigned to the same dose, and if two or more patients experience a DLT, the next cohort is treated at the next lower dose level, if possible. If already six patients have been observed at a dose level, the next cohort is treated at the higher dose level if less than two toxicities have been observed in the six patients. Otherwise, the next cohort is treated at the next lower dose level. The maximum number of patients treated at each dose level is six, and the trial is stopped, when we would either re-escalate to a dose where two or more out of three or six patients have experienced a DLT, or when we would de-escalate to a dose where we already have observed six patients. The MTD is then defined as the highest dose level where less than two out of six patients have experienced a DLT. That means the estimated MTD is the highest dose with observed toxicity rate less than $\frac{1}{3}$. Depending on the exact outcomes, this is the last dose administered or one dose below.

A graphical display of how to conduct a 3+3 design can be seen in Figure 2.2.

Different versions of the 3+3 design can be found in the literature, especially versions without de-escalation (e.g. in Ivanova (2006a)). Whenever the 3+3 design is mentioned, we refer to the version presented here.

The 3+3 design has some properties that make it popular for the use in practice. It is easy to understand and it can be followed very simply without complicated interim calculations. Therefore its application in a clinical study is quite convenient. The mathematical properties of this design are discussed in Section 6.1.



1

Figure 2.2: Flowchart for the 3+3 Design.

3 Introduction to Optimal Design Theory

In this chapter, we want to give an introduction to optimal design theory. The theory presented here is based on Silvey (1980), Fedorov (1972), Fedorov and Hackl (1997) and Atkinson and Donev (1996). An overview of the topic can also be found in the first chapters of Schwabe (1996).

The general goal of optimal experimental design is to determine experimental settings that maximize the amount of relevant information gained by the experiment.

We are interested in the functional dependence of a response variable Y on a set of r control variables $x_1, \ldots, x_r$, also denoted by the vector $\boldsymbol{x} \in \mathbb{R}$. Y is a random variable with a distribution given by a probability density function $p(y; \boldsymbol{x}, \boldsymbol{\theta})$. The vector of control variables $\boldsymbol{x}$ is not random and can be controlled by the experimenter. The variables $x_1, \ldots, x_r$ can be set to any values within the design space (also called design region), which is a given set $\mathcal{X} \subseteq \mathbb{R}^r$.

The relation between Y and $\boldsymbol{x}$ can be described as

$$E(Y(\boldsymbol{x})) = \eta(\boldsymbol{x}; \boldsymbol{\theta})$$

with $E(Y(\boldsymbol{x}))$ being the expected value of Y for given $\boldsymbol{x}$. The function $\eta(\boldsymbol{x}; \boldsymbol{\theta})$ depends on a set of p unknown parameters $\theta_1, \ldots, \theta_p$, denoted by the vector $\boldsymbol{\theta} \in \Theta$ with $\Theta \subset \mathbb{R}^p$.

We will confine us to the case that the function $\eta(\boldsymbol{x}; \boldsymbol{\theta})$ describing the relation of Y and $\boldsymbol{x}$ is known, and that only $\boldsymbol{\theta}$ or functions thereof are of interest.

Optimal design therefore has the goal of finding the experimental setting $\boldsymbol{x}$ that maximizes the precision of the estimated parameter vector $\hat{\boldsymbol{\theta}}$ or functions thereof, or analogously minimizes the variance $\mathrm{Cov}(\hat{\boldsymbol{\theta}})$ of the estimated parameters .

Throughout this chapter, we will first define what we mean by a design, then introduce information matrices and optimality criteria. We will also state some fundamental results of optimal design theory from the literature, that are very useful in the numerical search for optimal designs. Finally, we will introduce efficiencies as a measure for comparing different designs.

3.1 Definition of Continuous and Exact Designs

A continuous design ξ is a probability measure ξ on $\mathcal{X}$. It is only necessary to consider design measures with finite support, since only these are of practical relevance (cf. Kiefer (1959), page 281). Denote a design ξ by

$$\xi = \left\{ \begin{array}{ccc} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ w_1 & \cdots & w_n \end{array} \right\}$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are n distinct design points, that is different settings of the vector of control variables $\boldsymbol{x}$, and w_i are weights giving the proportion of the total observations taken at design point $\boldsymbol{x}_i$, where $0 \leq w_i \leq 1$ for all i and $\sum_{i=1}^{n} w_i = 1$. Let Ξ denote the set of all possible designs on $\mathcal{X}$.

Continuous designs are not always useful in practice, since the weights might give non-integer number of repetitions for some of the design points. A design measure that is realizable in integers for a given N is called an exact design or exact N-observation design and is denoted by

$$\xi_N = \left\{ \begin{array}{ccc} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ m_1/N & \cdots & m_n/N \end{array} \right\}$$

with $m_i \in \mathbb{N}, i = 1, ..., n$, and $\sum_{i=1}^{n} m_i = N$.

Numerical algorithms as will be presented in Section 3.5 aim at finding the optimal continuous design. If N is sufficiently large, good exact designs can usually be found by rounding of $w_i N$ to the nearest integer.

3.2 Information Matrices

To quantify the precision of the parameter estimates, the dispersion matrix, denoted by $\mathrm{Cov}(\hat{\boldsymbol{\theta}})$ is used.

In nonlinear models, the dispersion matrix is usually not available, but can be asymptotically approximated by the inverse of the Fisher information matrix.

The Fisher information matrix for a single observation on Y at $\boldsymbol{x}$ is given by the $p \times p$-dimensional matrix $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\theta})$ with elements

$$\boldsymbol{M}_{ij}(\boldsymbol{x}, \boldsymbol{\theta}) = E\left(\frac{\partial l}{\partial \theta_i} \frac{\partial l}{\partial \theta_j} \right) \ , \ i, j = 1, \ldots, p$$

where l denotes the log-likelihood function and is given by $l(\boldsymbol{\theta}\,; Y, \boldsymbol{x}) = \log p(Y\,; \boldsymbol{x}, \boldsymbol{\theta})$ (cf. Cox and Hinkley (2000)).

Assuming independent observations, Fisher information matrices are additive, and the overall information matrix for a design ξ is given by

$$\boldsymbol{M}(\xi, \boldsymbol{\theta}) = \sum_{i=1}^{n} w_i \boldsymbol{M}(x_i, \boldsymbol{\theta}).$$

Let $\hat{\boldsymbol{\theta}}$ be an unbiased estimator of $\boldsymbol{\theta}$. Then the diagonal elements of the inverse information matrix are lower bounds for the variances of the respective elements of $\hat{\boldsymbol{\theta}}$, i.e.

$$\mathrm{Var}(\hat{\theta}_i) \geq \boldsymbol{M}_{ii}^{-1}(\xi, \boldsymbol{\theta})$$

(Cramér-Rao inequality, see Cox and Hinkley (2000), page 256). It even holds that

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geq \boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta})$$

with respect to the Loewner order (cf. Witting (1985), page 317).

Under certain rather mild regularity conditions, like finite dimension of the parameter space and differentiability of the log likelihood (for details see Cox and Hinkley (2000), page 281), the inverse of the information matrix is the covariance matrix of the limiting distribution in case of asymptotic normality, i.e.

$$\boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta}) \overset{asymp.}{\approx} \mathrm{Cov}(\hat{\boldsymbol{\theta}}).$$

Therefore we will use the inverse of the Fisher information matrix as an approximation to the covariance matrix.

Let us here mention two properties of information matrices.

Theorem 3.1 (cf. Fedorov and Hackl (1997), Theorem 2.3.1)
Denote the set of all possible information matrices for a given $\boldsymbol{\theta}$ *by* $\mathcal{M}_{\boldsymbol{\theta}} = \{\boldsymbol{M}(\xi, \boldsymbol{\theta}); \xi \in \Xi\}$.

1. $\mathcal{M}_{\boldsymbol{\theta}}$ *is a convex set*
2. *all* $\boldsymbol{M}(\xi, \boldsymbol{\theta}) \in \mathcal{M}_{\boldsymbol{\theta}}$ *are symmetric and non-negative definite.*

This theorem is stated for linear models in Fedorov and Hackl (1997), but also holds for nonlinear models if $\boldsymbol{\theta}$ is fixed (cf. Silvey (1980)).

By Carathèodory's Theorem (see Appendix 2 in Silvey (1980)), any element of $\mathcal{M}_{\boldsymbol{\theta}}$ can be expressed as a weighted sum of at most $\frac{p(p+1)}{2}+1$ information matrices $\boldsymbol{M}(\xi_i, \boldsymbol{\theta})$ with ξ_i being the design putting unit weight at the design point $\boldsymbol{x}_i$. Thus there exists an optimal design containing $\frac{p(p+1)}{2}$ design points or less (cf. Atkinson and Donev (1996), page 96).

For nonlinear models, the information matrix $\boldsymbol{M}(\xi, \boldsymbol{\theta})$ depends on the unknown parameters $\boldsymbol{\theta}$. Thus it is not possible to generally optimize $\boldsymbol{M}(\xi, \boldsymbol{\theta})$ independent

of $\boldsymbol{\theta}$. There are different approaches of dealing with the dependence on the unknown parameters. Here we will focus on locally optimal designs, which are designs optimizing - in some sense - the information matrix for a given value of $\boldsymbol{\theta}$.

3.3 Optimality Criteria

We will use the information matrix as a measure for the amount of information gained by the experiment. The aim is to maximize this information. We will now specify in which way we can maximize the information by using the information matrix.

A very strong conditions for a design ξ^* to maximize the information gained and thus to be optimal would be that $\boldsymbol{M}(\xi^*, \boldsymbol{\theta}) - \boldsymbol{M}(\xi, \boldsymbol{\theta})$ is non-negative definite for all $\xi \in \Xi$. Such a 'uniformly optimal design' usually does not exist.

Therefore we will focus on maximizing real-valued functions of the information matrix. Depending on the goal of the experiment, different functions are useful. These functions ψ are referred to as optimality criteria. A design ξ^* is called locally ψ-optimal at $\boldsymbol{\theta}$, if

$$\xi^* = \arg\max_{\xi} \psi(M(\xi, \boldsymbol{\theta})).$$

In this section, we want to present some of the most important and commonly used optimality criteria.

3.3.1 The D-Criterion

A D-optimal design ξ_D is a design that maximizes the determinant of the information matrix, i.e.

$$\xi_D = \arg\max_{\xi} \det \boldsymbol{M}(\xi, \boldsymbol{\theta}).$$

The D-optimal design is not necessarily unique.

For computational convenience, often $-\log\det \boldsymbol{M}(\xi, \boldsymbol{\theta})$ is minimized, which is an equivalent optimization problem.

The D-criterion determines the design that asymptotically minimizes the volume of a confidence ellipsoid of the parameter vector. Although often the volume of the confidence ellipsoid has no practical meaning, the D-criterion is commonly used due to its favorable mathematical properties.

The D-criterion is for example invariant to linear transformations of the model (cf. Atkinson and Donev (1996), page 117). The D-optimal design is independent of the actual parametrization used in the model (cf. Pázman (1986), page 79).

3.3.2 The c-Criterion

The c-criterion is used to minimize the variance of a linear function $\boldsymbol{c}^T\boldsymbol{\theta}$ of the parameters.

The c-optimal design ξ_c is given by

$$\xi_c = \arg\max_{\xi} \left(\boldsymbol{c}^T \boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta})\boldsymbol{c}\right)^{-1} = \arg\min_{\xi} \boldsymbol{c}^T \boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta})\boldsymbol{c}$$

where $\boldsymbol{c}$ is the vector giving the linear combination of $\boldsymbol{\theta}$ of interest. Problems can arise when the information matrix of the optimal design is singular. Therefore we will only admit designs that allow parameter estimation. For those designs, the information matrix is non-singular, but we have to note that the c-optimal design might not exist on this restricted set of designs as it is not a closed and compact set.

Special cases of the c-criterion can be used, when the goal is to minimize the variance of a single parameter θ_i. Then $\boldsymbol{c}$ is a vector with entries only 0, and 1 for the i-th entry. Analogously $\boldsymbol{c}$ can be chosen to maximize the precision of the estimate of a certain quantile of the distribution of Y.

3.3.3 The L-Criterion

When several linear combinations of the elements of the parameter vector $\boldsymbol{\theta}$ are of interest, the following linear criterion is applicable. The design ξ_L is L-optimal if

$$\xi_L = \arg\min_{\xi} \operatorname{tr}\left(\boldsymbol{L}^T \boldsymbol{M}(\xi, \boldsymbol{\theta})^{-1} \boldsymbol{L}\right)$$

where $\boldsymbol{L}$ is a matrix whose columns correspond to the different linear combinations of the elements of $\boldsymbol{\theta}$ that are of interest. Thus ξ_L is the design that minimizes the sum of the variances of the estimates of those linear combinations. The same problems related to singular information matrices as mentioned for the c-criterion can occur. Again, we will only consider designs with non-singular information matrices, being aware that the optimal design for this restricted design region might not exist.

3.4 The General Equivalence Theorem

In this section we will quote some theoretical results that are fundamental for optimal designs theory and vital for the numerical search for optimal designs.

First let us introduce the directional derivative.

Definition 3.2 *The Fréchet derivative of ψ at $\boldsymbol{M}_1$ in the direction of $\boldsymbol{M}_2$ is defined as*

$$F_\psi(\boldsymbol{M}_1, \boldsymbol{M}_2) = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left(\psi((1-\epsilon)\boldsymbol{M}_1 + \epsilon \boldsymbol{M}_2) - \psi(\boldsymbol{M}_1)\right).$$

Now we can state the following.

Theorem 3.3 (cf. Silvey (1980), Theorem 3.6)
If ψ is convex on $\mathcal{M}_{\boldsymbol{\theta}}$, ξ^* is ψ-optimal if and only if

$$F_\psi(\boldsymbol{M}(\xi^*, \boldsymbol{\theta}), \boldsymbol{M}(\xi, \boldsymbol{\theta})) \geq 0 \quad \textit{for all } \xi \in \Xi.$$

That means a design is optimal, if it cannot be improved by moving slightly in the direction of any other design. In the case of differentiability of ψ, it suffices to check whether the design is improved by changing it in the direction of any one-point design measure.

Theorem 3.4 (cf. Silvey (1980), Theorem 3.7)
If ψ is convex on $\mathcal{M}_{\boldsymbol{\theta}}$ and differentiable at $\boldsymbol{M}(\xi^*)$, ξ^* is ψ-optimal if and only if

$$F_\psi(\boldsymbol{M}(\xi^*, \boldsymbol{\theta}), \boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\theta})) \geq 0 \quad \textit{for all } \boldsymbol{x} \in \mathcal{X}.$$

This theorem is of greater practical use than the one above.

We will now define the sensitivity function, which is used in the general equivalence theorem by Kiefer and Wolfowitz.

Definition 3.5 *The sensitivity function d of ξ at $\boldsymbol{x} \in \mathcal{X}$ is given by*

$$d(\boldsymbol{x}, \xi, \boldsymbol{\theta}) = tr\left(\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\theta}) \boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta})\right).$$

Theorem 3.6 (cf. Kiefer and Wolfowitz (1960))
The following statements are equivalent: the design ξ_D is D-optimal if

1. $\xi_D = arg\min_\xi(-\log det \boldsymbol{M}(\xi, \boldsymbol{\theta}))$,
2. $\xi_D = arg\min_\xi \max_{\boldsymbol{x}} d(\boldsymbol{x}, \xi, \boldsymbol{\theta})$,
3. $\max_{\boldsymbol{x}} d(\boldsymbol{x}, \xi_D, \boldsymbol{\theta}) = p$, *where p is the dimension of $\boldsymbol{\theta}$.*

This is one of the most important results in optimal design theory and of great practical relevance. The equivalence theorem can be generalized to other optimality criteria.

A representation of the sensitivity function and the equivalence theorem directly applicable to the cases we will consider is given in Dragalin et al. (2006) (formula 16): A design ξ_D is locally D-optimal if and only if

$$d(\boldsymbol{x}, \xi_D, \boldsymbol{\theta}) = \operatorname{tr}\left(\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\theta}) \boldsymbol{M}^{-1}(\xi_D, \boldsymbol{\theta})\right) \leq p, \tag{3.1}$$

for all $\boldsymbol{x} \in \mathcal{X}$ and $d(\boldsymbol{x}, \xi_D, \boldsymbol{\theta}) = p$ at all support points of ξ_D. p is the total number of parameters.

3.5 Numerical Methods and Algorithms for the Construction of Optimal Designs

The optimization problems for finding optimal designs usually cannot be solved analytically. Numerical algorithms based on the statements of the general equivalence theorem have been developed to help find the solution to the optimization problems. The algorithm used for the construction of D-optimal designs in Chapters 4 and 5 is Fedorov's First Order Algorithm for D-optimality as described in Chapter 3.1 in Fedorov and Hackl (1997).

This algorithm is based on the fact that the sensitivity function $d(\boldsymbol{x}, \xi_D, \boldsymbol{\theta})$ achieves its maxima at the support points of the D-optimal design.

We will start with an arbitrary design ξ_1 with non-singular information matrix $\boldsymbol{M}(\xi_1, \boldsymbol{\theta})$. At each iteration step s, we aim at improving the design ξ_s by putting more weight on the point where $d(\boldsymbol{x}, \xi_s, \boldsymbol{\theta})$ is maximal. So we find $\boldsymbol{x}_s = \arg\max_x d(\boldsymbol{x}, \xi_s, \boldsymbol{\theta})$ and add the point $\boldsymbol{x}_s$ to the design. That means we construct

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s \xi(\boldsymbol{x}_s)$$

with $\xi(\boldsymbol{x}_s)$ being the unit measure at x_s. The value for $\alpha_s \in (0, 1)$ is chosen to fulfill

$$\alpha_s = \arg\max_{\alpha} \det\left(\boldsymbol{M}((1 - \alpha)\xi_s + \alpha\xi(\boldsymbol{x}_s), \boldsymbol{\theta})\right),$$

so we choose the proportion α of the new design point to maximize the gain in the optimality criterion.

3.6 Efficiency of Designs

When comparing different design, the efficiency is a useful measure. We consider two types of efficiencies.

Definition 3.7 *The D-efficiency of a design ξ is defined as*

$$D_{eff}(\xi) = \left(\frac{det\boldsymbol{M}(\xi, \boldsymbol{\theta})}{det\boldsymbol{M}(\xi_D, \boldsymbol{\theta})}\right)^{\frac{1}{p}},$$

where ξ_D is the D-optimal design and p the total number of parameters.

Definition 3.8 *The G-efficiency of a design ξ is defined as*

$$G_{eff}(\xi) = \frac{p}{\max_{\boldsymbol{x} \in \mathcal{X}} d(\boldsymbol{x}, \xi, \boldsymbol{\theta})},$$

where p is the total number of parameters.

The D-efficiency represents the amount of information the design under investigation ξ yields, as compared to the D-optimal design. E.g., having a design with D-efficiency of 0.5, it would have to be repeated twice to yield the same precision of the estimates as the D-optimal design. The interpretation of the G-efficiency is analogous with respect to the G-optimal design minimizing $\max_{\boldsymbol{x}\in\mathcal{X}} d(\boldsymbol{x}, \xi, \boldsymbol{\theta})$. The D-efficiency can only be determined if the D-optimal design and thus the value of the D-criterion for the optimal design is known. However the G-efficiency can be calculated even if the G-optimal design is not known, and therefore can be applied more easily in practice. The G-efficiency of the D-optimal design is 1. This follows directly from the general equivalence theorem.

For any design $\xi \in \Xi$, it holds that

$$G_{eff} \leq D_{eff}$$

(cf. Corollary 3 in Dette (1996)). Thus the easy to determine G-efficiency can be used as a lower bound for the D-efficiency.

4 Models used for Dose Finding Studies

Analyzing dose finding studies as presented in Chapter 2 might require some specific methods. When modelling the dose response relationship, the ordinary linear model often is not sufficient.

Depending on the type of response – binary, categorical or continuous – different models have to be used. Additionally, the dose response relationship is commonly assumed to be non-linear and often monotonically increasing. These assumptions have to be taken into account as well when choosing the model for the dose response relationship.

In this chapter, we want to present some models appropriate for analyzing dose response relationships. Additionally we give an overview over optimal designs for the respective models where relevant information is available.

4.1 The Logistic Model

The ordinary 2-parameter logistic model as presented in Hosmer and Lemeshov (1989) or Agresti (1990) is often used for modelling binary response variables. The expected response is not modelled directly, instead the probability for a certain outcome is modelled. The logistic model belongs to the class of generalized linear models.

4.1.1 Definition of the Model

Let Y be the binary response with outcomes denoted by 0 and 1, 0 meaning failure and 1 meaning success. Success in this case means observing the event of interest. $\boldsymbol{x} \in \mathbb{R}^p$ is the vector of control variables, possibly including an intercept. The ordinary logistic model is then defined as follows.

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

with $\boldsymbol{\beta} \in \mathbb{R}^p$ being the parameter vector. $P(Y(\boldsymbol{x}) = 1)$ is the probability of observing a response of 1 given the value $\boldsymbol{x}$ of the control variables.

For the case of an intercept and a single control variable, this model reduces to

$$P(Y(x) = 1) = \frac{\exp(\alpha + \beta \cdot x)}{1 + \exp(\alpha + \beta \cdot x)}.$$

With this model we get a relationship between the outcome and the control variable that is monotonic in x. Additionally, $0 \leq P(Y(x) = 1) \leq 1$, thus this is a reasonable way of modelling probabilities.

In literature dealing with medical applications (e.g. Ting (2006a)), a different definition of the model is quite common. Let Y again be the response variable and x the single control variable, as above. The logistic regression function is then defined as

$$P(Y(x) = 1) = \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{1 + \exp\left(\frac{x-\mu}{\sigma}\right)}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. The parameter μ corresponds to the value of x for which the probability of the response being 1 is 0.5, i.e. $P(Y(\mu) = 1) = 0.5$. In medical applications the control variable often is the dose and μ is often referred to as the ED_{50}, 'ED' meaning 'effective dose'. Thus μ is the dose showing an effect of the drug in half the population, or in terms of probability, having a probability of 0.5 that the drug shows an effect. The second parameter σ is related to the steepness of the slope. It does not have a practical meaning as μ does, but smaller values of σ lead to a steeper curve. Restricting σ to positive values gives a function that is strictly monotonically increasing in x.

This representation of the model is favored in medical applications due to the direct interpretability of the parameter μ. Thus we will refer to this representation as the 'medical parametrization' of the logistic model, as opposed to the 'classical parametrization' as given above. If the parameter β in the classical parametrization of the model is restricted to positive values, both representations of the model are equivalent, and $\mu = -\frac{\alpha}{\beta}$ and $\sigma = \frac{1}{\beta}$.

Whenever we refer to the logistic model within this work, it will be the one with a single control variable and the medical parametrization, unless stated otherwise.

4.1.2 Parameter Estimation

Estimation of the parameters in the logistic model is done using the maximum likelihood approach. The existence of the maximum likelihood estimator (MLE)

though cannot be guaranteed, unless the observations fulfill some specific conditions. We need observations in both of the response categories, and the observations (x_i, y_i) must not be separable. That means there has to be an overlap of the values of x for which $Y = 0$ and $Y = 1$. The conditions can be found in a more formal representation in Silvapulle (1981).

For the classical parametrization, the conditions of part (III) of the theorem in Silvapulle (1981) are satisfied and the theorem provides a necessary and sufficient condition for the existence of the maximum likelihood estimator. This condition can be rephrased as follows.

Theorem 4.1 (cf. Silvapulle (1981))
Let $\mathcal{X}_0 = \{x_i | y_i = 0\}$ *and* $\mathcal{X}_1 = \{x_i | y_i = 1\}$. *The maximum likelihood estimate in the ordinary logistic model in the classical parametrization with one control variable exists and is unique if and only if*

$$(\min(\mathcal{X}_0) < \max(\mathcal{X}_1)) \wedge (\min(\mathcal{X}_1) < \max(\mathcal{X}_0)).$$

The following examples will illustrate these conditions.

Example 4.2 *Consider three design points* $x_1 < x_2 < x_3$. *We have one observation at each design point. The following observations are possible:*

	x_1	x_2	x_3
$\boldsymbol{y}_1$	0	0	0
$\boldsymbol{y}_2$	0	0	1
$\boldsymbol{y}_3$	**0**	**1**	**0**
$\boldsymbol{y}_4$	0	1	1
$\boldsymbol{y}_5$	1	0	0
$\boldsymbol{y}_6$	**1**	**0**	**1**
$\boldsymbol{y}_7$	1	1	0
$\boldsymbol{y}_8$	1	1	1

Only for two of these eight possible outcomes (marked in boldface), the conditions of Theorem 4.1 are met and the MLE exists.

Example 4.3 *We have again three design points* $x_1 < x_2 < x_3$, *but now we have two observation at each design point. Consider two possible outcomes:*

	x_1	x_2	x_3
$\boldsymbol{y}_1$	0, 0	0, 1	0, 1
$\boldsymbol{y}_2$	0, 0	0, 1	1, 1

In the first case, the conditions in Theorem 4.1 are met and thus the MLE exists, since

$$\begin{aligned} \min(\mathcal{X}_0) = x_1 &< x_3 = \max(\mathcal{X}_1) \quad \textit{and} \\ \min(\mathcal{X}_1) = x_2 &< x_3 = \max(\mathcal{X}_0). \end{aligned}$$

In the second case, the conditions are not met, since

$$\min(\mathcal{X}_1) = x_2 = \max(\mathcal{X}_0).$$

Thus in this case, the MLE does not exist.

To state conditions for the existence of the MLE in the medical parametrization, we take a closer look at the parameter β in the classically parameterized model. If the MLE exists in this model, and $\hat{\beta} > 0$, the MLE in the medical parametrization is given by $\hat{\mu} = -\frac{\hat{\alpha}}{\hat{\beta}}$ and $\hat{\sigma} = \frac{1}{\hat{\beta}}$.

Let us first state the following lemma.

Lemma 4.4 *Let $g(x)$ be a real-valued strictly monotonical or constant function and $x_i \in \mathbb{R}, i = 1, \ldots, n$, with $x_i \leq x_{i+1} \, \forall \, i$. Let there exist at least one i such that $x_i < x_{i+1}$. Define $\overline{g} = \frac{1}{n} \sum_{i=1}^{n} g(x_i)$. Then*

1. $\sum_{i=1}^{n} x_i g(x_i) = \sum_{i=1}^{n} x_i \overline{g} \;\Leftrightarrow\; g$ *is constant,*
2. $\sum_{i=1}^{n} x_i g(x_i) > \sum_{i=1}^{n} x_i \overline{g} \;\Leftrightarrow\; g$ *is strictly monotonically increasing,*
3. $\sum_{i=1}^{n} x_i g(x_i) < \sum_{i=1}^{n} x_i \overline{g} \;\Leftrightarrow\; g$ *is strictly monotonically decreasing.*

Proof: 1. Let us first show that $\sum_{i=1}^{n} x_i g(x_i) = \sum_{i=1}^{n} x_i \overline{g} \;\Rightarrow\; g$ is constant.

Assume g is strictly monotonically increasing. Then

$$\exists l \text{ with } x_l < x_{l+1} : \forall i \leq l \;:\; g(x_i) < \overline{g} \quad \wedge \quad \forall i > l \;:\; g(x_i) \geq \overline{g}.$$

Note that

$$\sum_{i=1}^{l} (\overline{g} - g(x_i)) = \sum_{i=l+1}^{n} (g(x_i) - \overline{g}) > 0.$$

By the initial condition we have

$$\begin{aligned} \sum_{i=1}^{n} x_i g(x_i) &= \sum_{i=1}^{n} x_i \overline{g} \\ \Leftrightarrow \sum_{i=1}^{n} x_i (g(x_i) - \overline{g}) &= 0 \\ \Leftrightarrow \sum_{i=l+1}^{n} x_i (g(x_i) - \overline{g}) &= \sum_{i=1}^{l} x_i (\overline{g} - g(x_i)) \end{aligned}$$

By replacing x_i on the left hand side of the inequation with its minimum possible value x_{l+1}, we decrease the term on this side. Analogously we increase the term on the right hand side by replacing x_i with its maximum possible value x_l. Thus it follows that

$$\begin{aligned} \sum_{i=l+1}^{n} x_{l+1}(g(x_i)-\overline{g}) &\leq \sum_{i=1}^{l} x_l(\overline{g}-g(x_i)) \\ \Leftrightarrow x_{l+1}\sum_{i=l+1}^{n}(g(x_i)-\overline{g}) &\leq x_l\sum_{i=1}^{l}(\overline{g}-g(x_i)) \\ \Leftrightarrow x_{l+1} &\leq x_l. \end{aligned}$$

This is a contradiction to the assumption. Analogously a contradiction can be shown for the assumption that g is strictly monotonically decreasing. Thus we can conclude that g has to be constant.

We can conclude equivalence since the other direction, i.e.

$$g \text{ is constant } \Rightarrow \sum_{i=1}^{n} x_i g(x_i) = \sum_{i=1}^{n} x_i \overline{g}$$

is obvious.

2. Let us first show that

$$g \text{ is strictly monotonically increasing } \Rightarrow \sum_{i=1}^{n} x_i g(x_i) > \sum_{i=1}^{n} x_i \overline{g}.$$

As a condition, we have $x_i \leq x_{i+1}, i = 1, ..., n$. Thus for g being strictly monotonically increasing, it follows that

$$g(x_i) \leq g(x_{i+1}).$$

By Chebyshev's sum inequality (cf. Hardy et al. (1988), page 43-44), we can conclude directly that

$$\begin{aligned} \sum_{i=1}^{n} x_i g(x_i) &\geq \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} g(x_i) \\ \Leftrightarrow \sum_{i=1}^{n} x_i g(x_i) &\geq \sum_{i=1}^{n} x_i \overline{g}. \end{aligned}$$

Additionally, we can rule out equality because of part 1 of this lemma.

3. Analogously we can show that

$$g \text{ is strictly monotonically decreasing } \Rightarrow \sum_{i=1}^{n} x_i g(x_i) < \sum_{i=1}^{n} x_i \overline{g}.$$

It still remains to be shown that

$$\begin{aligned} \sum_{i=1}^{n} x_i g(x_i) > \sum_{i=1}^{n} x_i \overline{g} &\Rightarrow g \text{ is strictly monotonically increasing,} \\ \sum_{i=1}^{n} x_i g(x_i) < \sum_{i=1}^{n} x_i \overline{g} &\Rightarrow g \text{ is strictly monotonically decreasing.} \end{aligned}$$

This follows logically since the three cases considered are disjoint and include all possible cases. □

Now we can state the following theorem.

Theorem 4.5
The MLE for the logistic model in the medical parametrization exists if and only if

1. $(\min(\mathcal{X}_0) < \max(\mathcal{X}_1)) \wedge (\min(\mathcal{X}_1) < \max(\mathcal{X}_0))$
2. $\frac{1}{r}\sum_{x_i \in \mathcal{X}_1} x_i > \frac{1}{n-r}\sum_{x_i \in \mathcal{X}_0} x_i$ *with* $r = \sum_{i=1}^{n} y_i$.

Proof: The first condition follows directly from Theorem 4.1 and the relationship between the parameters in the different representations of the model. The second condition ensures that the maximum likelihood estimate for β in the classical parametrization is positive. Define $g(x) = \frac{\exp(\alpha+\beta\cdot x)}{1+\exp(\alpha+\beta\cdot x)}$. Then $g(x)$ is strictly monotonically increasing for $\beta > 0$, strictly monotonically decreasing for $\beta < 0$ and constant for $\beta = 0$.

The negative log-likelihood for the classically parameterized model is given by

$$l(\alpha, \beta) = -r\alpha - \beta \sum_{x_i \in \mathcal{X}_1} x_i + \sum_{i=1}^{n} \ln(1 + \exp(\alpha + \beta x_i)).$$

It is convex in the parameters (see Pratt (1981)), thus any extremum has to be a global minimum. The negative log-likelihood $l(\alpha, \beta)$ achieves its minimum if

$$\begin{aligned} \frac{\partial l}{\partial \alpha} &= -r + \sum_{i=1}^{n} \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} = 0 \qquad (4.1) \\ \wedge \quad \frac{\partial l}{\partial \beta} &= -\sum_{x_i \in \mathcal{X}_1} x_i + \sum_{i=1}^{n} x_i \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} = 0. \end{aligned}$$

Thus it has to be

$$\begin{aligned} \sum_{i=1}^{n} x_i g(x_i) &= \sum_{i \in \mathcal{X}_1} x_i \\ \Leftrightarrow \frac{1}{r}\sum_{i=1}^{n} x_i g(x_i) &= \frac{1}{r}\sum_{x_i \in \mathcal{X}_1} x_i. \end{aligned}$$

With the second condition stated in this theorem, it follows that

$$\frac{1}{r}\sum_{x_i \in \mathcal{X}_1} x_i > \frac{1}{n-r}\sum_{x_i \in \mathcal{X}_0} x_i \quad \Rightarrow \frac{1}{r}\sum_{x_i \in \mathcal{X}_1} x_i > \frac{1}{n}\sum_{i=1}^{n} x_i$$

and thus

$$\begin{aligned} \frac{1}{r}\sum_{i=1}^{n} x_i g(x_i) &> \frac{1}{n}\sum_{i=1}^{n} x_i \\ \Leftrightarrow \sum_{i=1}^{n} x_i g(x_i) &> \sum_{i=1}^{n} x_i \frac{r}{n}. \end{aligned}$$

By equation 4.1, we can replace r with $\sum_{i=1}^{n} g(x_i)$ and we get

$$\sum_{i=1}^{n} x_i g(x_i) > \sum_{i=1}^{n} x_i \frac{\sum_{i=1}^{n} g(x_i)}{n}.$$

Define $\frac{1}{n}\sum_{i=1}^{n} g(x_i) =: \overline{g}$ and we get

$$\sum_{i=1}^{n} x_i g(x_i) > \sum_{i=1}^{n} x_i \overline{g}.$$

With Lemma 4.4, it follows that $g(x)$ is monotonically increasing and thus $\hat{\beta} > 0$.

Analogously, it can be shown that if $\frac{1}{r}\sum_{x_i \in \mathcal{X}_1} x_i < \frac{1}{n-r}\sum_{x_i \in \mathcal{X}_0} x_i \Rightarrow \hat{\beta} < 0$ and if $\frac{1}{r}\sum_{x_i \in \mathcal{X}_1} x_i = \frac{1}{n-r}\sum_{x_i \in \mathcal{X}_0} x_i \Rightarrow \hat{\beta} = 0$. Since these three cases are disjoint and include all possible cases, we can conclude equivalence. □

This theorem offers a convenient way to check if parameter estimation is possible with the given observations. It also allows for estimating the probability of being able to determine the MLE, given a design and assumed parameters.

4.1.3 Optimal Design for the Logistic Model

Optimal design theory for the logistic model is well developed (see for example Silvey (1980), page 60). The information matrix is given by

$$M(\xi, \boldsymbol{\theta}) = \sum_{i=1}^{n} w_i M(x_i, \boldsymbol{\theta}) = \sum_{i=1}^{n} w_i \frac{1}{\sigma^2} F(x_i, \boldsymbol{\theta})(1 - F(x_i, \boldsymbol{\theta})) \begin{pmatrix} 1 & \frac{x_i - \mu}{\sigma} \\ \frac{x_i - \mu}{\sigma} & \left(\frac{x_i - \mu}{\sigma}\right)^2 \end{pmatrix},$$

where

$$\boldsymbol{\theta} = \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \text{ and } F(x_i, \boldsymbol{\theta}) = \frac{\exp\left(\frac{x_i - \mu}{\sigma}\right)}{1 + \exp\left(\frac{x_i - \mu}{\sigma}\right)}.$$

It can be shown that the design

$$\xi = \begin{pmatrix} \mu - 1.5434\sigma & \mu + 1.5434\sigma \\ 0.5 & 0.5 \end{pmatrix}$$

is D-optimal for the logistic model with parameters μ and σ. This can easily be verified numerically using Equation 3.1.

The design points of this D-optimal design correspond to the 0.176 and 0.824 quantiles of the logistic distribution function. The equal weights indicate that exact designs with equal number of replications at both design points are D-optimal. The design points of the D-optimal design are illustrated in Figure 4.1.

The D-optimal design allows for maximum likelihood estimation of the parameters, if both possible responses are observed at both design points, and if the

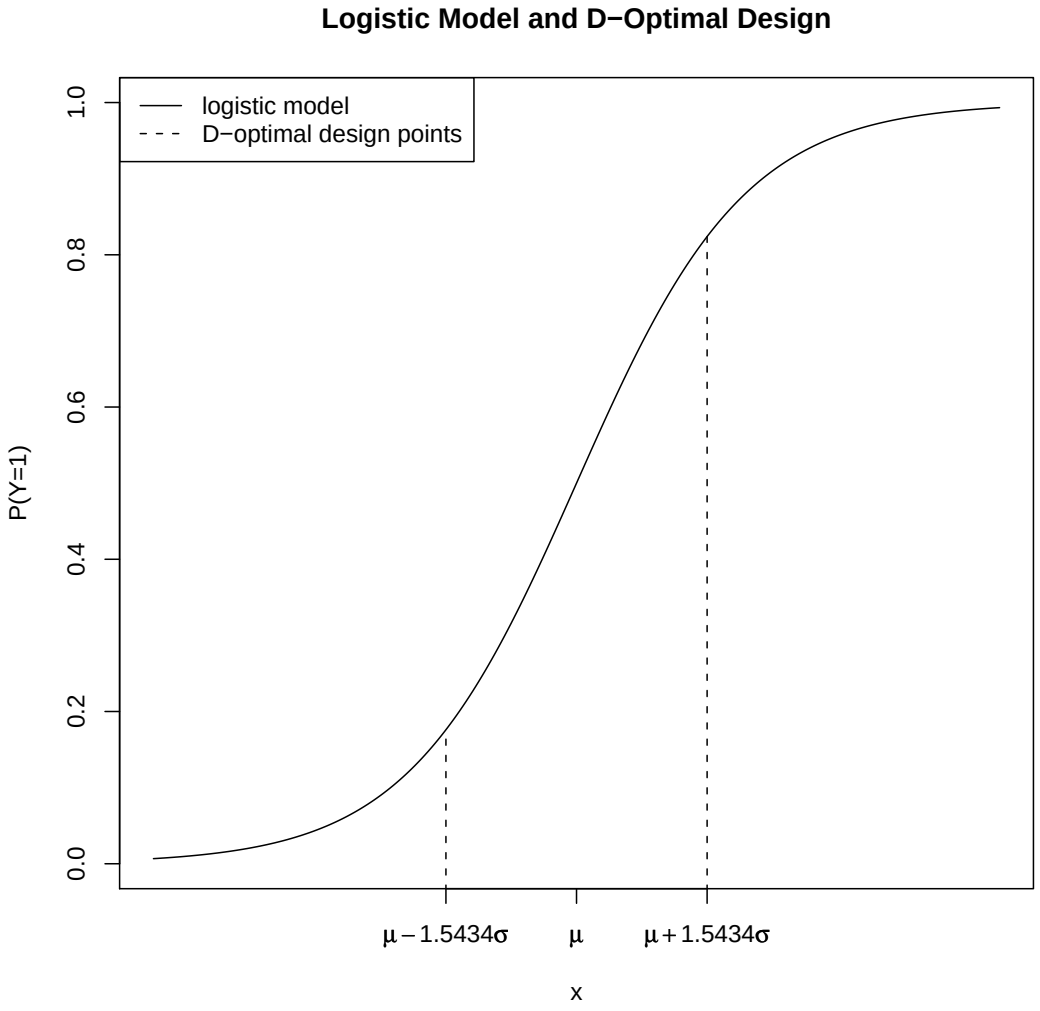


Figure 4.1: Example of a logistic model and corresponding D-optimal design points.

number of responses equal to 1 is greater at the larger design point than at the smaller one. This follows directly from Theorem 4.5. So from a practical point of view, to fulfill these conditions for estimability, at least three replications are needed at each of the two design points.

4.2 The Proportional Odds Model

The proportional odds model as described in McCullagh (1980) and in Agresti (1990) is a generalization of the 2-parameter logistic model to an ordinal response variable with more than two categories.

4.2.1 Definition of the Model

Let Y be the response variable and x a control variable. Assume Y is observed in $K+1$ ordered categories $j = 0, \ldots, K$, then the cumulative probability of $Y \geq j$ can be modelled as

$$P(Y(x) \geq j) = \frac{\exp(\frac{x-\alpha_j}{\beta})}{1+\exp(\frac{x-\alpha_j}{\beta})}, j = 1, \ldots, K$$

where $\beta > 0$ and $\alpha_1 < \ldots < \alpha_K$.

The parameters α_j give the values of x for which $P(Y(\alpha_j) \geq j) = 0.5$. Thus these parameters can be interpreted analogously as the ED_{50} in the ordinary logistic model.

The common slope parameter β is needed to assure valid cumulative probabilities. If we would allow a possibly different β_j for each category, we might get $P(Y(x) \geq j) \geq P(Y(x) \geq j+1)$, and thus a negative probability for $P(Y(x) = j)$. This is not admissible, and can be avoided by setting $\beta_j = \beta$ for all j.

The probability of the outcome being j is given by

$$\begin{aligned} p_j := P(Y(x) = j) &= \begin{cases} 1 - P(Y(x) \geq 1), & j = 0 \\ P(Y(x) \geq j) - P(Y(x) \geq j+1), & j = 1, \ldots, K-1 \\ P(Y(x) \geq K), & j = K \end{cases} \\ &= \begin{cases} \frac{1}{1+\exp(\frac{x-\alpha_1}{\beta})}, & j = 0 \\ \frac{\exp(\frac{x-\alpha_j}{\beta})}{1+\exp(\frac{x-\alpha_j}{\beta})} - \frac{\exp(\frac{x-\alpha_{j+1}}{\beta})}{1+\exp(\frac{x-\alpha,j+1}{\beta})}, & j = 1, \ldots, K-1 \\ \frac{\exp(\frac{x-\alpha_K}{\beta})}{1+\exp(\frac{x-\alpha_K}{\beta})}, & j = K \end{cases} \end{aligned}$$

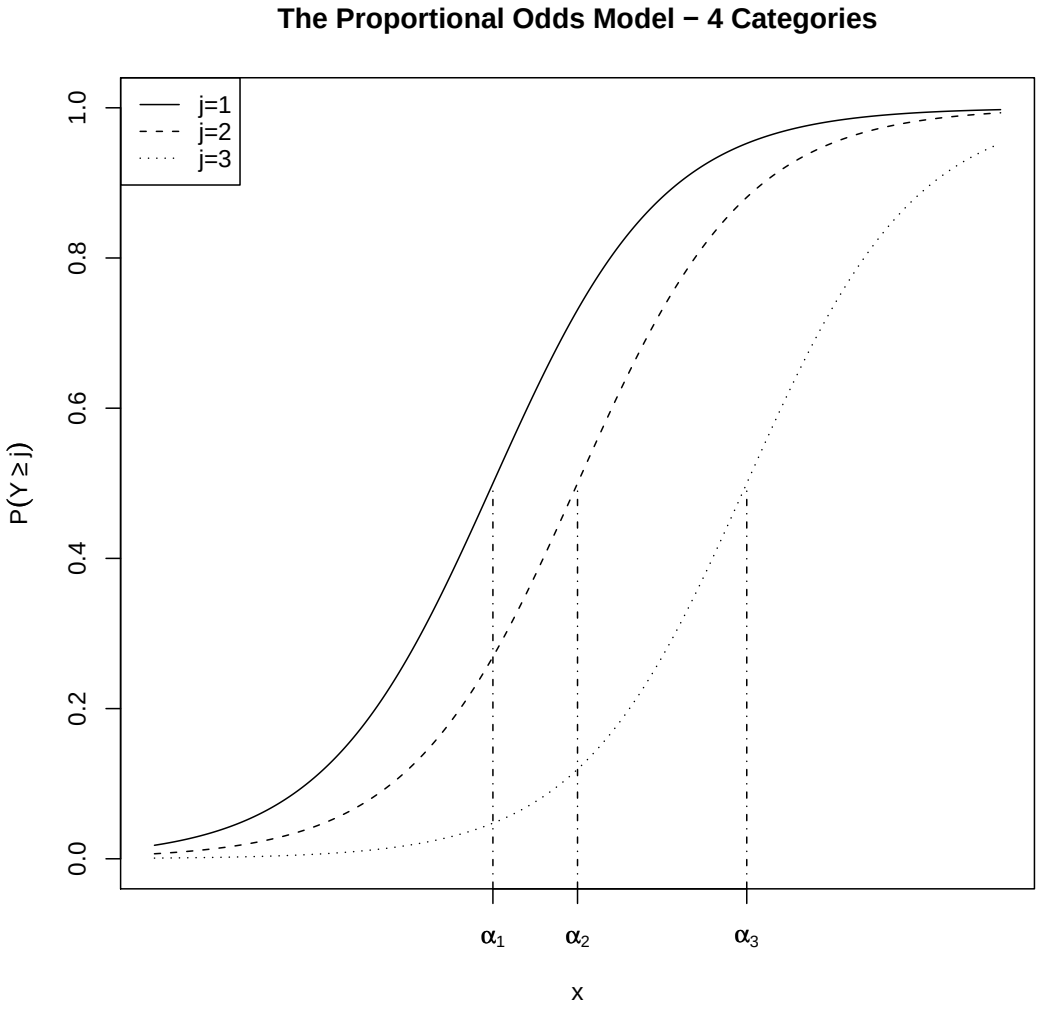


Figure 4.2: Example for a proportional odds model with 4 categories.

An example for a proportional odds model with four categories is given in Figure 4.2.

4.2.2 Properties of the Model

The use of this model can be motivated by the following. Consider an underlying continuous response variable Z, that can only be observed in $K+1$ categories. Let Z have a cumulative distribution function $F_Z(z) = 1 - G(\frac{x-z}{\beta})$. Let $Y = j$ if

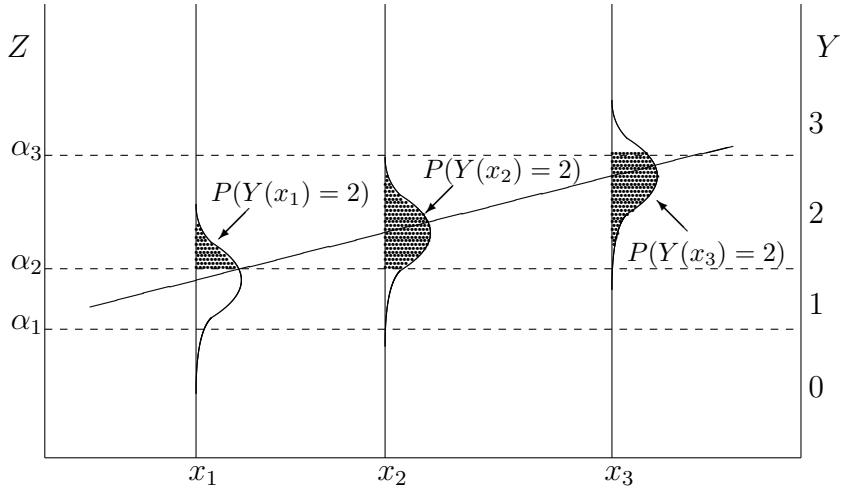


Figure 4.3: Underlying continuous regression model and ordinal measurement, cf. Agresti (1990), Figure 9.2.

$\alpha_j < Z \leq \alpha_{j+1}$ where $-\infty = \alpha_0 < \ldots < \alpha_{K+1} = \infty$. Then

$$P(Y(x) \geq j) = P(Z(x) > \alpha_j) = 1 - P(Z(x) \leq \alpha_j) = G\left(\frac{x - \alpha_j}{\beta}\right).$$

Choosing G as the inverse logit function, we get the model defined above (cf. Agresti (1990)). A graphical interpretation is given in Figure 4.3

We will now state another property of this model. Consider the cumulative odds ratio of $Y(x_1) \leq j$ and $Y(x_2) \leq j$, that is

$$\frac{P(Y(x_1) \leq j) \;/\; P(Y(x_1) > j)}{P(Y(x_2) \leq j) \;/\; P(Y(x_2) > j)} = \exp\left(\frac{x_1 - x_2}{\beta}\right).$$

Thus the log of the cumulative odds ratio is proportional to the difference in the control variable, independent of the category considered. The name 'proportional odds model' is ascribed to this property.

If $K = 1$, the response is binary and the proportional odds model reduces to the ordinary logistic model. Therefore for the remainder of this text, the logistic model will be included in the proportional odds model as a special case.

4.2.3 Parameter Estimation

Concerning the existence of the maximum likelihood estimator, similar problems occur as in the ordinary logistic models. Observations in all of the $K+1$ categories are necessary for the maximum likelihood estimator to exist. Additionally, again a certain degree of overlap between the observation points with observations in the

different categories is necessary. The exact conditions are specified in Habermann (1980), and they are rephrased in Liu et al. (2009), where the proof can be found in Liu (2006). These conditions only apply to a linear parametrization of the model, but still give a necessary condition for the model as defined above.

4.2.4 Optimal Design for the Proportional Odds Model

Locally optimal designs for the proportional odds model are derived in Perevozskaya et al. (2003). The information matrix for a single design point for this model is given by

$$\boldsymbol{M}(x, \boldsymbol{\theta}) = \frac{1}{\beta^2} \begin{bmatrix} \mathbb{I}_K \\ \boldsymbol{v}^T \end{bmatrix} \boldsymbol{D}\boldsymbol{P}\boldsymbol{D} \begin{bmatrix} \mathbb{I}_K & \boldsymbol{v} \end{bmatrix},$$

with $\mathbb{I}_K$ being the identity matrix of dimension K and

$$\boldsymbol{D} = \mathrm{diag}_{j=1,\ldots,K} \left(P(Y(x) \geq j)(1 - P(Y(x) \geq j)) \right),$$

$$\boldsymbol{v} = \begin{pmatrix} \frac{x-\alpha_1}{\beta} \\ \vdots \\ \frac{x-\alpha_K}{\beta} \end{pmatrix} \text{ and}$$

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{p_0} + \frac{1}{p_1} & -\frac{1}{p_1} & 0 & \cdots & 0 \\ -\frac{1}{p_1} & \frac{1}{p_1} + \frac{1}{p_2} & -\frac{1}{p_2} & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & \ddots & 0 \\ \vdots & & & & -\frac{1}{p_{K-1}} \\ 0 & \cdots & 0 & -\frac{1}{p_{K-1}} & \frac{1}{p_{K-1}} + \frac{1}{p_K} \end{bmatrix}.$$

The information matrix for a single design point thus is of dimension $K+1$, but of rank K. Therefore at least two different design points are necessary for having a nonsingular overall information matrix.

Locally optimal designs for this model can be determined using an iterative algorithm, e.g. the one presented in Section 3.5.

It suffices to consider standardized parameters (cf. Lemma 2 in Perevozskaya et al. (2003)). They are chosen to be $\beta = 1$ and either $\bar{\alpha} = \frac{1}{K} \sum_{j=1}^{K} \alpha_j = 0$ or $\alpha_j = 0$ for one $j, j \in \{1, \ldots, K\}$.

For the case of four categories, i.e. $K = 3$, and standardized parameters $\beta = 1, \alpha_2 = 0$, the optimal design is shown in Figure 4.4. We only consider the special case of equidistant categories such that $\alpha_1 - \alpha_2 = \alpha_2 - \alpha_3$ and with the

standardized $\alpha_2 = 0$ we get $\alpha_1 = -\alpha_3$. In the upper frame, the optimal design points are shown. The number of support points for the optimal design varies between two and six. The according weights are shown in the lower frame. The optimal design only depends on the difference between α_j and α_{j+1}, $j = 1, 2$. For large differences between α_1, α_2 and α_3, the optimal design converges to a six-point design with equal weights and the design points being approximately $\alpha_j \pm 1.043$.

4.3 The $\mathbf{E}_{\max}$ Model

The $E_{\max}$ model as presented in MacDougall (2006) is a common model used in analyzing dose response relationships. It can be used for both continuous and binary outcomes and is quite flexible.

4.3.1 Definition of the Model

Let Y be the response variable and x a control variable. Then the $E_{\max}$ model is defined as follows:

$$E(Y(x)) = E_0 + \frac{x^\lambda \cdot E_{\max}}{x^\lambda + ED_{50}^\lambda}.$$

E_0 is the expected response for $x = 0$, $E_{\max}$ the maximum effect of x on Y, ED_{50} the value of x that yields half the maximum effect and λ a slope factor related to the steepness of the curve. If the response is a continuous outcome, the parameters E_0 and $E_{\max}$ do not have to be restricted.

If the response variable is binary, the $E_{\max}$ model can also be used by modelling $P(Y(x))$ as

$$P(Y(x) = 1) = E_0 + \frac{x^\lambda \cdot E_{\max}}{x^\lambda + ED_{50}^\lambda}$$

with $0 \leq E_0 \leq 1$ and $0 \leq E_0 + E_{\max} \leq 1$ to ensure a valid response on the probability scale.

The function given by the $E_{\max}$ model is monotonic. It is decreasing, if $E_{\max}$ is negative, and increasing, if $E_{\max}$ is positive.

Some examples for the $E_{\max}$ model with different parameters can be found in Figure 4.5.

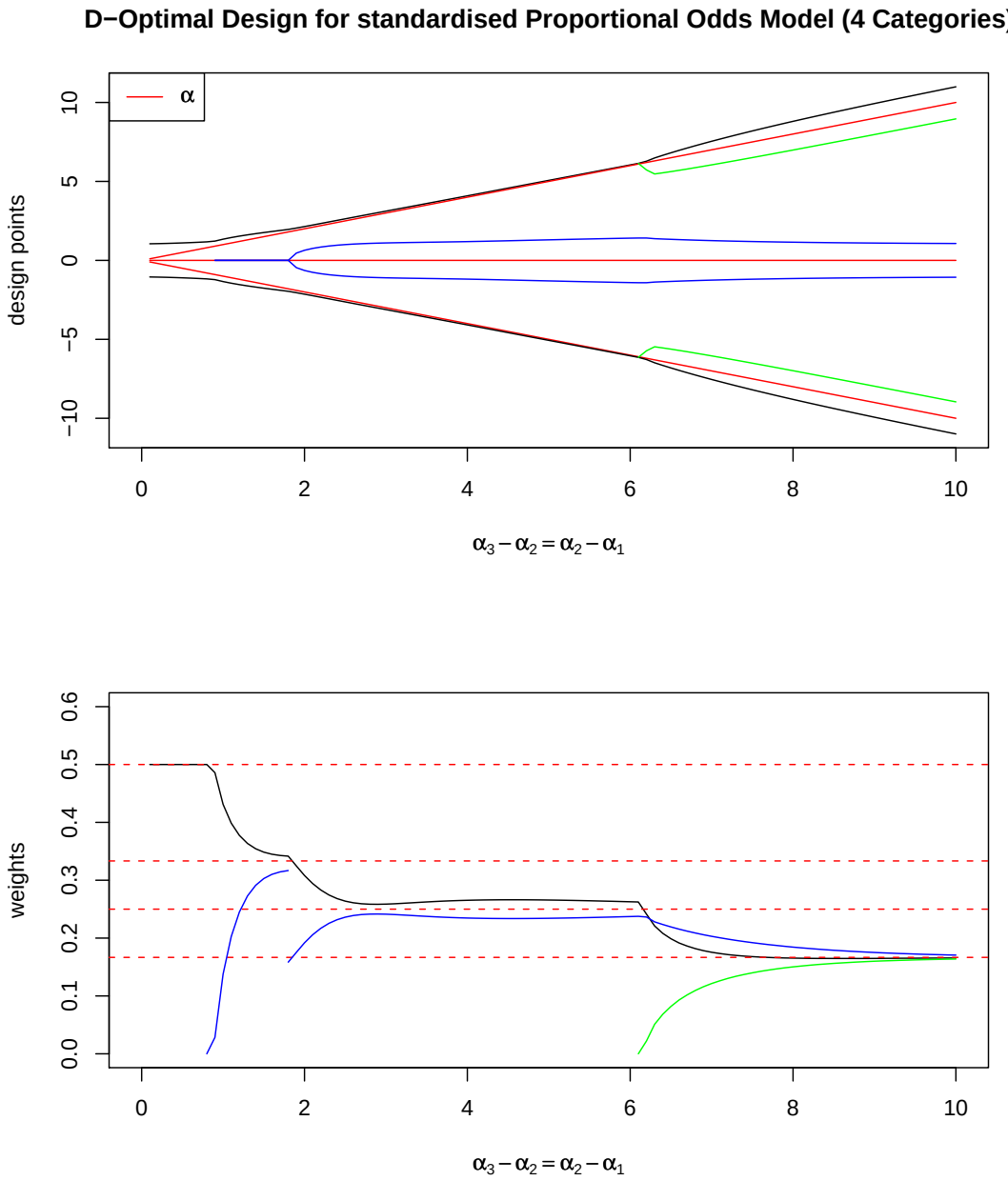


Figure 4.4: D-optimal design for a standardized 4-category proportional odds model; top: optimal design points, bottom: optimal weights.

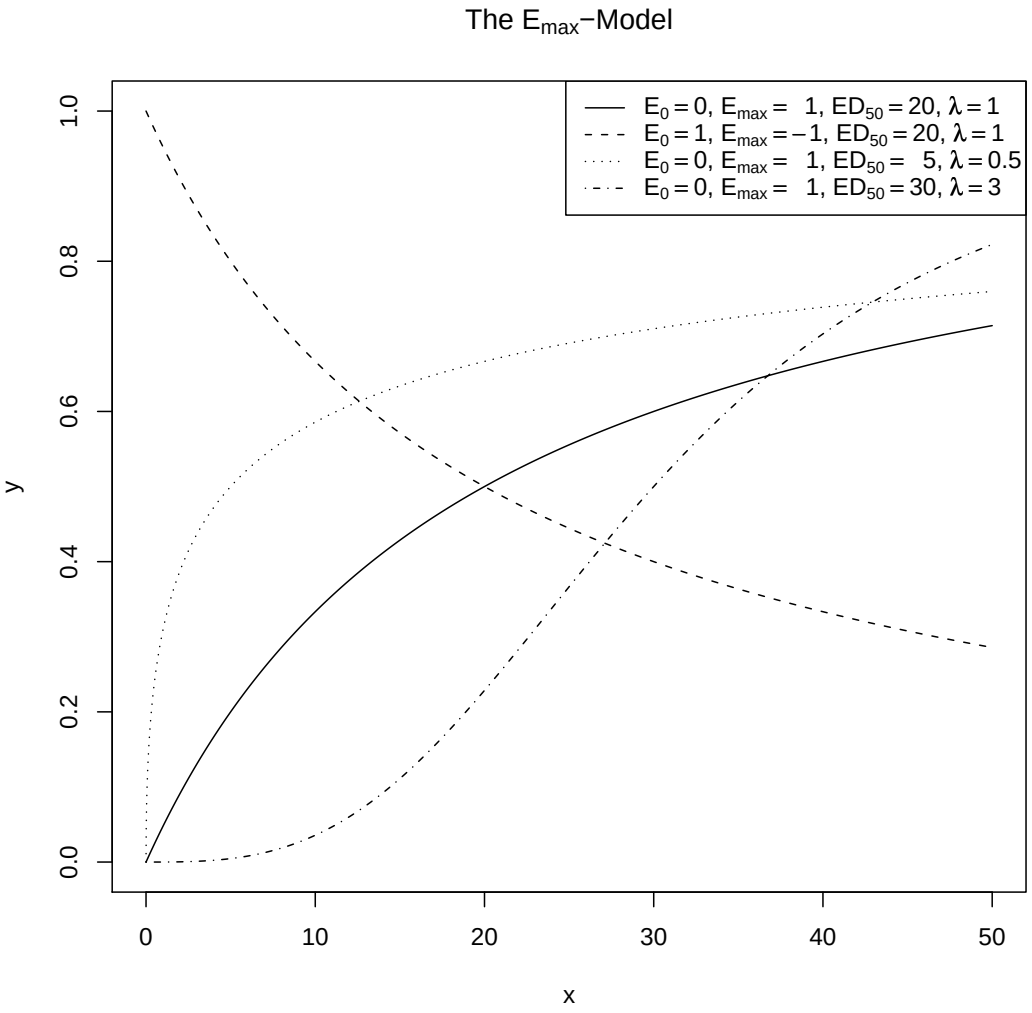


Figure 4.5: Example of the $E_{\max}$ model for different parameters.

4.3.2 Properties of the Model

The $E_{\max}$ model for binary outcomes is closely related to the logistic model. It can be interpreted as a logistic model on the log scale of the control variable. It can be easily seen that the above model can be rewritten as

$$E(y) = E_0 + \frac{E_{\max}}{1 + \exp\left(-\lambda(\log(x) - \log(ED_{50}))\right)}.$$

With $E_0 = 0$ and $E_{\max} = 1$ and the setting for a binary response variable, we have

$$P(Y(x) = 1) = \frac{\exp\left(\frac{\log(x) - \log(ED_{50})}{\lambda^{-1}}\right)}{1 + \exp\left(\frac{\log(x) - \log(ED_{50})}{\lambda^{-1}}\right)}.$$

This corresponds to the ordinary logistic model with $\log(x)$ as the control variable instead of x, $\log(ED_{50})$ corresponding to μ and λ^{-1} being the equivalent of σ.

When E_0 is not restricted to the value 0 and $E_{\max}$ to the value 1, the $E_{\max}$ model is more flexible than the logistic model. This is due to the two additional parameters, that relate to the range of the outcome Y or $P(Y(x) = 1)$, respectively. However, due to the larger number of parameters, observations at more distinct designs points are needed for estimating these. The 4-parameter $E_{\max}$ model thus is not desirable if only very few design points are available.

Optimal design for the $E_{\max}$ model is not of interest for this work and therefore shall not be discussed. The model though is needed for comparisons in Chapter 7.

5 A Bivariate Model for Safety and Efficacy

The models and associated designs described in Chapter 4 consider only one response variable (in medical applications commonly called endpoint), i.e. either efficacy or toxicity of the target drug, although often studies are conducted to gather information on both endpoints.

When two endpoints should be considered simultaneously, a bivariate model is needed. Dragalin and Fedorov (2006) and Dragalin et al. (2006) suggest some models to consider if both endpoints are measured on a binary scale and they construct locally optimal designs for these models. Here, a model is introduced that allows one endpoint to be measured on a categorical scale, while the other is measured on a binary scale. Subsequently, the properties of the model are described and locally optimal designs are derived.

5.1 Definition of the Model

Consider a bivariate response variable $\boldsymbol{Y} = (T\ ,\ E)^T$, with T being the toxicity endpoint and E the efficacy endpoint. Without loss of generality let the efficacy endpoint be measured on a binary scale with possible outcomes 0 (no efficacy) and 1 (efficacy), while the toxicity endpoint is observed in K+1 categories $j = 0, 1, ..., K$, where the higher category indicates stronger toxicity. These categories can e.g. be defined by the Common Terminology Criteria for Adverse Events (cf. National Cancer Institute (2006)).

Consider a single control variable x, namely the dose. The notation for the probabilities for each of the possible bivariate outcomes and the marginal probabilities are shown in Table 5.1.

Here, $P(T(x) = y_T)$ and $P(E(x) = y_E)$ denote the probability of the outcome being $y_T \in \{0, \ldots, K\}$ and $y_E \in \{0, 1\}$ given the treatment dose x.

As described in Chapter 4, the logistic model and the proportional odds model are reasonable models for binary and ordered categorical outcomes, respectively. Therefore it is desirable that the bivariate modelling is analogous. To achieve this, the marginal distributions of the considered endpoints should follow a logistic and a proportional odds model.

	E		
T	0	1	Σ
0	p_{00}	p_{01}	$p_{0.} = 1 - P(T(x) \geq 1)$
1	p_{10}	p_{11}	$p_{1.} = P(T(x) \geq 1) - P(T(x) \geq 2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	p_{j0}	p_{j1}	$p_{j.} = P(T(x) \geq j) - P(T(x) \geq j+1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
K	p_{K0}	p_{K1}	$p_{K.} = P(T(x) \geq K)$
Σ	$p_{.0} =$ $1 - P(E(x) = 1)$	$p_{.1} =$ $P(E(x) = 1)$	1

Table 5.1: Probabilities for the outcomes in the bivariate model.

Define $F(x) := \frac{exp(x)}{1+exp(x)}$ and let the marginal distribution of the efficacy endpoint be given by

$$P(E(x) = 1) = \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{1 + \exp\left(\frac{x-\mu}{\sigma}\right)} = F\left(\frac{x-\mu}{\sigma}\right).$$

For notational convenience let $x_\mu := \frac{x-\mu}{\sigma}$ and thus $P(E(x) = 1) = F(x_\mu)$.

The marginal distribution of the toxicity endpoint is given by

$$P(T(x) \geq j) = \frac{\exp\left(\frac{x-\alpha_j}{\beta}\right)}{1 + \exp\left(\frac{x-\alpha_j}{\beta}\right)} = F\left(\frac{x-\alpha_j}{\beta}\right).$$

Here let $x_{\alpha_j} := \frac{x-\alpha_j}{\beta}, \quad \alpha = 1, ..., K$ and thus $P(T(x) \geq j) = F(x_{\alpha_j})$. This gives consistency with adequate univariate modelling.

A joint distribution that yields the above marginal distributions and that we will use for the bivariate modelling is given by the following functions.

Define $G(x, y) := F(x)F(y)\{1 + \tau[1 - F(x)][1 - F(y)]\}$. Then

$$\begin{aligned} P(T(x) \geq j \wedge E(x) = 1) &= F(x_{\alpha_j})F(x_\mu)\{1 + \tau\left[1 - F(x_{\alpha_j})\right][1 - F(x_\mu)]\} \\ &= G(x_{\alpha_j}, x_\mu). \end{aligned}$$

This is a bivariate distribution function from the class of Farlie-Gumbel-Morgenstern distributions (cf. Kotz et al. (2000), Chapter 44.13), which arises quite naturally from the given univariate marginal distributions.

Thus we have a model with K+4 parameters denoted by the vector $\boldsymbol{\theta} = (\mu, \alpha_1, ..., \alpha_K, \sigma, \beta, \tau)^T$ where $\mu \in \mathbb{R}$, $-\infty < \alpha_1 < ... < \alpha_K < \infty$, $\sigma, \beta > 0$ and $-1 \leq \tau \leq 1$. Note that for $\tau = 0$, $T(x)$ and $E(x)$ are independent.

Define $G(\infty, \cdot) := \lim_{x\to\infty} G(x, \cdot)$ and $G(\cdot, \infty)$ analogously as the corresponding limit, and $\alpha_0 = -\infty$ and $\alpha_{K+1} = \infty$.

It can be easily seen that the marginal probabilities are

$$\begin{aligned}
p_{.0} &= 1 - G(\infty, x_\mu) = 1 - F(x_\mu) \\
p_{.1} &= G(\infty, x_\mu) = F(x_\mu) \\
p_{0.} &= 1 - G(x_{\alpha_1}, \infty) = 1 - F(x_{\alpha_1}) \\
p_{j.} &= G(x_{\alpha_j}, \infty) - G(x_{\alpha_{j+1}}, \infty) = F(x_{\alpha_j}) - F(x_{\alpha_{j+1}}), \quad j = 1, ..., K-1 \\
p_{K.} &= G(x_{\alpha_K}, \infty) = F(x_{\alpha_K}).
\end{aligned}$$

The joint probabilities are given by

$$\begin{aligned}
p_{01} &= G(\infty, x_\mu) - G(x_{\alpha_1}, x_\mu) = F(x_\mu) - G(x_{\alpha_1}, x_\mu) \\
p_{j1} &= G(x_{\alpha_j}, x_\mu) - G(x_{\alpha_{j+1}}, x_\mu), \quad j = 1, ..., K-1 \\
p_{K1} &= G(x_{\alpha_K}, x_\mu) \\
p_{j0} &= p_{j.} - p_{j1}, \quad j = 0, ..., K, \text{ so} \\
p_{00} &= 1 - F(x_{\alpha_1}) - F(x_\mu) + G(x_{\alpha_1}, x_\mu) \\
p_{j0} &= F(x_{\alpha_j}) - F(x_{\alpha_{j+1}}) - G(x_{\alpha_j}, x_\mu) + G(x_{\alpha_{j+1}}, x_\mu), \quad j = 1, ..., K-1 \\
p_{K0} &= F(x_{\alpha_K}) - G(x_{\alpha_K}, x_\mu).
\end{aligned}$$

Since $F(x_{\alpha_0}) = 1$, $F(x_{\alpha_{K+1}}) = 0$, $G(x_{\alpha_0}, x_\mu) = F(x_\mu)$ and $G(x_{\alpha_{K+1}}, x_\mu) = 0$, the joint probabilities can be written as

$$\begin{aligned}
p_{j0} &= F(x_{\alpha_j}) - F(x_{\alpha_{j+1}}) - G(x_{\alpha_j}, x_\mu) + G(x_{\alpha_{j+1}}, x_\mu), \quad j = 0, ..., K \text{ and} \\
p_{j1} &= G(x_{\alpha_j}, x_\mu) - G(x_{\alpha_{j+1}}, x_\mu), \quad j = 0, ..., K.
\end{aligned}$$

5.2 Properties of the Model

This model describes the relationship between dose and efficacy, and dose and each of the toxicity categories, respectively. The relationship is such that the probability for efficacy and toxicity of a certain grade, respectively, is monotonically increasing with the dose.

The probability for any grade of toxicity and efficacy reaches 1 for infinitely large doses, and it is larger than 0 for a dose of 0. More precisely,

$$\begin{aligned}
\lim_{x \to \infty} P(T(x) \geq j \text{ and } E(x) = 1) = 1 \text{ and} \\
P(T(0) \geq j \text{ and } E(0) = 1) > 0
\end{aligned}$$

for all $1 \leq j \leq K$.

Since the toxicity categories are ordered, the modelling of cumulative probabilities as done here is appropriate.

The dependence between the two endpoints is modelled by the parameter τ. For $\tau = 0$, both endpoints are independent, for $\tau > 0$ we have a positive correlation, and for $\tau < 0$ the correlation is negative.

An example for the bivariate model with $K = 3, \mu = 5, \sigma = 20, \alpha_1 = 16, \alpha_2 = 22, \alpha_3 = 30$ and $\beta = 8$ is shown in Figures 5.1 and 5.2. Figure 5.1 displays the joint probabilities. The upper frame shows the case of independence (i.e. $\tau = 0$), whereas in the lower frame $\tau = 0.8$ and thus the case of positive correlation between both endpoints is shown. The marginal probabilities for efficacy and the toxicity categories, which do not depend on τ, are displayed in Figure 5.2.

5.3 Optimal Designs for this Model

Following common design theory (cf. Chapter 3), optimal designs are constructed by maximizing real-valued functions of the information matrix in order to maximize the information obtained by the experiment.

As we are dealing with a non-linear model, the information matrix depends on the unknown parameters, and we will focus on locally optimal designs.

5.3.1 Information Matrices

The information matrix of a single observation at design point x given the parameter vector $\boldsymbol{\theta} = (\mu, \alpha_1, ..., \alpha_K, \sigma, \beta, \tau)^T$ is denoted by $\boldsymbol{M}(x, \boldsymbol{\theta})$, and can be derived from

$$\boldsymbol{M}(x, \boldsymbol{\theta}) = E\left(\frac{\partial l}{\partial \boldsymbol{\theta}} \frac{\partial l}{\partial \boldsymbol{\theta}^T}\right)$$

where l denotes the log-likelihood function of a single observation $\boldsymbol{y} = (j, i)$. It is given by

$$l(\boldsymbol{\theta}; x, \boldsymbol{y}) = \log P(T(x) = j \wedge E(x) = i) I_{\{T=j, E=i\}} = \log(p_{ji}) I_{\{T=j, E=i\}}.$$

Thus

$$\begin{aligned}
\boldsymbol{M}(x, \boldsymbol{\theta}) &= E\left(\frac{\partial \log(p_{ji}) I_{\{T=j,E=i\}}}{\partial \boldsymbol{\theta}} \frac{\partial \log(p_{ji}) I_{\{T=j,E=i\}}}{\partial \boldsymbol{\theta}^T}\right) \\
&= E\left(\frac{1}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \boldsymbol{\theta}} \frac{\partial p_{ji}}{\partial \boldsymbol{\theta}^T} I_{\{T=j,E=i\}}\right) \\
&= \sum_{i,j} \frac{1}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \boldsymbol{\theta}} \frac{\partial p_{ji}}{\partial \boldsymbol{\theta}^T} p_{ji} \\
&= \sum_{i,j} \frac{1}{p_{ji}} \frac{\partial p_{ji}}{\partial \boldsymbol{\theta}} \frac{\partial p_{ji}}{\partial \boldsymbol{\theta}^T}.
\end{aligned}$$

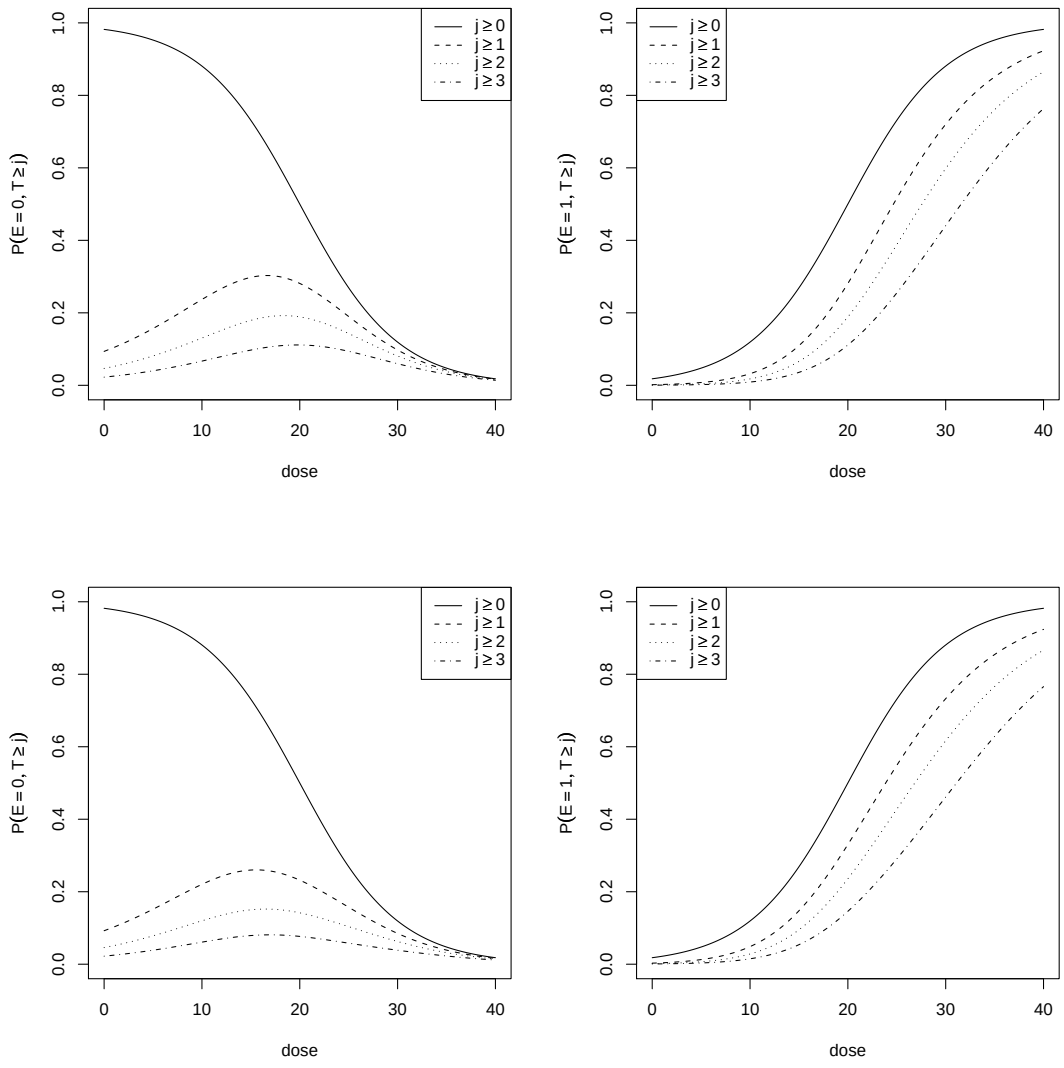


Figure 5.1: Joint probabilities in the bivariate model, top: $\tau = 0$, bottom: $\tau = 0.8$.

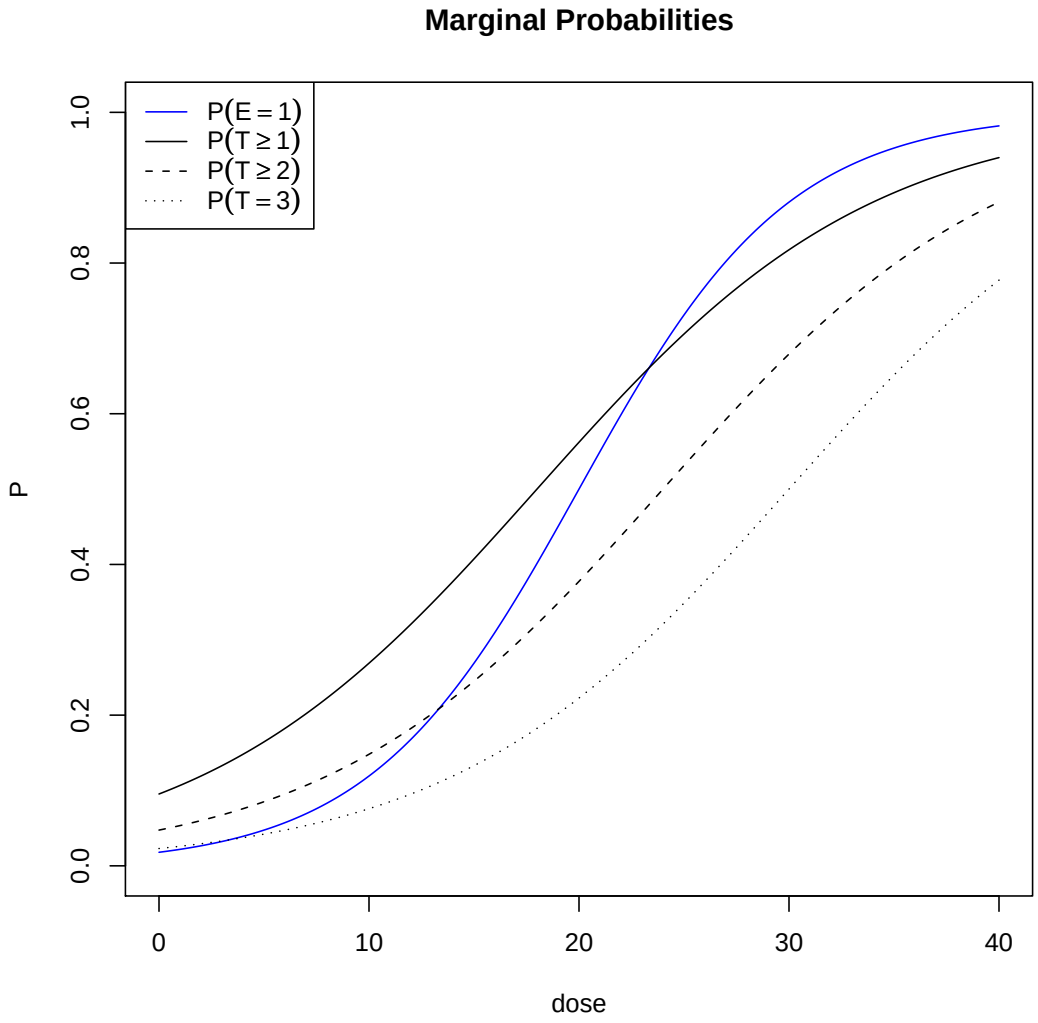


Figure 5.2: Marginal probabilities in the bivariate model.

The information matrix of a design ξ with design points x_i and corresponding weights w_i is given by $\boldsymbol{M}(\xi, \boldsymbol{\theta}) = \sum_i w_i \boldsymbol{M}(x_i, \boldsymbol{\theta})$.

Denote the full information matrix for a single observation by

$$\boldsymbol{M}(x, \boldsymbol{\theta}) = \left(\begin{array}{c|ccc|c|c|c} \boldsymbol{M}_{\mu\mu} & \boldsymbol{M}_{\mu\alpha_1} & \cdots & \boldsymbol{M}_{\mu\alpha_K} & \boldsymbol{M}_{\mu\sigma} & \boldsymbol{M}_{\mu\beta} & \boldsymbol{M}_{\mu\tau} \\ \hline \boldsymbol{M}_{\mu\alpha_1} & \boldsymbol{M}_{\alpha_1\alpha_1} & \cdots & \boldsymbol{M}_{\alpha_1\alpha_K} & \boldsymbol{M}_{\alpha_1\sigma} & \boldsymbol{M}_{\alpha_1\beta} & \boldsymbol{M}_{\alpha_1\tau} \\ \boldsymbol{M}_{\mu\alpha_2} & \boldsymbol{M}_{\alpha_1\alpha_2} & \cdots & \boldsymbol{M}_{\alpha_2\alpha_K} & \boldsymbol{M}_{\alpha_2\sigma} & \boldsymbol{M}_{\alpha_2\beta} & \boldsymbol{M}_{\alpha_2\tau} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{M}_{\mu\alpha_K} & \boldsymbol{M}_{\alpha_1\alpha_K} & \cdots & \boldsymbol{M}_{\alpha_K\alpha_K} & \boldsymbol{M}_{\alpha_K\sigma} & \boldsymbol{M}_{\alpha_K\beta} & \boldsymbol{M}_{\alpha_K\tau} \\ \hline \boldsymbol{M}_{\mu\sigma} & \boldsymbol{M}_{\alpha_1\sigma} & \cdots & \boldsymbol{M}_{\alpha_K\sigma} & \boldsymbol{M}_{\sigma\sigma} & \boldsymbol{M}_{\sigma\beta} & \boldsymbol{M}_{\sigma\tau} \\ \hline \boldsymbol{M}_{\mu\beta} & \boldsymbol{M}_{\alpha_1\beta} & \cdots & \boldsymbol{M}_{\alpha_K\beta} & \boldsymbol{M}_{\sigma\beta} & \boldsymbol{M}_{\beta\beta} & \boldsymbol{M}_{\beta\tau} \\ \hline \boldsymbol{M}_{\mu\tau} & \boldsymbol{M}_{\alpha_1\tau} & \cdots & \boldsymbol{M}_{\alpha_K\tau} & \boldsymbol{M}_{\sigma\tau} & \boldsymbol{M}_{\beta\tau} & \boldsymbol{M}_{\tau\tau} \end{array}\right).$$

The indices indicate which derivatives are taken to obtain the specific element of $\boldsymbol{M}$ and this notation allows for conveniently referring to single elements or submatrices of $\boldsymbol{M}$.

For notational convenience, let us define

$$H(x, y) := F(x)(1 + \tau(1 - F(x))(1 - 2F(y))).$$

Note that $H(x_{\alpha_0}, x_\mu) = 1$ and $H(x_{\alpha_{K+1}}, x_\mu) = 0$.

In the subsequent section we will only display the resulting information matrices, the derivation of these matrices can be found in Appendix A.

Partial Information Matrix

Let us first assume that the correlation parameter τ and the slope parameters σ and β are known and fixed. Then the information matrix reduces to the submatrix

$$\boldsymbol{M}\left(x, \begin{pmatrix} \mu \\ \alpha \end{pmatrix}\right) := \begin{pmatrix} \boldsymbol{M}_{\mu\mu} & \boldsymbol{M}_{\mu\alpha_1} & \dots & \boldsymbol{M}_{\mu\alpha_K} \\ \boldsymbol{M}_{\mu\alpha_1} & \boldsymbol{M}_{\alpha_1\alpha_1} & \dots & \boldsymbol{M}_{\alpha_1\alpha_K} \\ \vdots & & & \vdots \\ \boldsymbol{M}_{\mu\alpha_K} & & \dots & \boldsymbol{M}_{\alpha_K\alpha_K} \end{pmatrix}.$$

It is given by $\boldsymbol{M}\left(x, \begin{pmatrix} \mu \\ \alpha \end{pmatrix}\right) = \boldsymbol{D}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T\boldsymbol{D} \sim (K+1) \times (K+1)$ where

$$
\begin{aligned}
\boldsymbol{D} &= \operatorname{diag}\left(\frac{1}{\sigma}F(x_\mu)(1-F(x_\mu)), \frac{1}{\beta}F(x_{\alpha_1})(1-F(x_{\alpha_1})), \ldots, \frac{1}{\beta}F(x_{\alpha_K})(1-F(x_{\alpha_K}))\right) \\
&\quad \sim (K+1)\times(K+1), \\
\boldsymbol{P} &= \begin{pmatrix} \operatorname{diag}\left(\frac{1}{p_{j0}}\right)_{j=0}^{K} & 0 \\ 0 & \operatorname{diag}\left(\frac{1}{p_{j1}}\right)_{j=0}^{K} \end{pmatrix} \sim (2K+2)\times(2K+2), \\
\boldsymbol{H} &= (\boldsymbol{H}_1\ \boldsymbol{H}_2) \sim (K+1)\times(2K+2).
\end{aligned}
$$

The matrix $\boldsymbol{H}$ is composed of two similar looking matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, both of dimension $(K+1)\times(K+1)$. $\boldsymbol{H}_2$ is given by

$$
\boldsymbol{H}_2 = \left(\begin{array}{ccccccc}
\multicolumn{7}{c}{\boldsymbol{h}^T} \\ \hline
-H(x_\mu,x_{\alpha_1}) & H(x_\mu,x_{\alpha_1}) & 0 & & \cdots & & 0 \\
0 & -H(x_\mu,x_{\alpha_2}) & H(x_\mu,x_{\alpha_2}) & 0 & \cdots & & 0 \\
\vdots & & \ddots & & & & \vdots \\
 & & & & & & 0 \\
0 & \cdots & 0 & & -H(x_\mu,x_{\alpha_K}) & & H(x_\mu,x_{\alpha_K})
\end{array}\right)
$$

where

$$
\boldsymbol{h} = \begin{pmatrix} H(x_{\alpha_0},x_\mu) - H(x_{\alpha_1},x_\mu) \\ H(x_{\alpha_1},x_\mu) - H(x_{\alpha_2},x_\mu) \\ \vdots \\ H(x_{\alpha_K},x_\mu) - H(x_{\alpha_{K+1}},x_\mu) \end{pmatrix}.
$$

$\boldsymbol{H}_1$ has the same structure and is given by

$$
\boldsymbol{H}_1 = \left(\begin{array}{cccccc}
\multicolumn{6}{c}{\mathbf{0}} \\ \hline
-1 & 1 & 0 & & \cdots & 0 \\
0 & -1 & 1 & 0 & \cdots & 0 \\
\vdots & & \ddots & & & \vdots \\
 & & & & & 0 \\
0 & \cdots & 0 & & -1 & 1
\end{array}\right) - \boldsymbol{H}_2.
$$

Lemma 5.1 *The rank of* $\boldsymbol{M}\left(x, \begin{pmatrix} \mu \\ \alpha \end{pmatrix}\right)$ *is K+1, i.e.* $\boldsymbol{M}\left(x, \begin{pmatrix} \mu \\ \alpha \end{pmatrix}\right)$ *is of full rank.*

Proof: The matrices $\boldsymbol{D} \sim (K+1)\times(K+1)$ and $\boldsymbol{P} \sim (K+1)\times(K+1)$ are diagonal matrices with all diagonal elements being non-zero, and therefore are of full rank. Since

multiplication with full rank matrices does not affect the rank, $\boldsymbol{M}_{\mu\alpha}$ is of full rank if $\boldsymbol{H}$ is of full row rank and thus $\boldsymbol{H}^T$ is of full column rank. To show that $\boldsymbol{H} \sim (K+1)\times(2K+2)$ is of full rank, we need to show that rank$\boldsymbol{H} = K+1$. Consider only one part of $\boldsymbol{H}$, namely $\boldsymbol{H}_1$. $\boldsymbol{H}_1$ is of dimension $(K+1)\times(K+1)$.

Note that $0 < H(x,y) < 1$ for all $x, y \in \mathbf{R}$ (excluding $\pm\infty$), and that with h_i being the i-th element of $\boldsymbol{h}$, we have $\sum_{i=1}^{K+1} h_i = 1$.

If we multiply $\boldsymbol{H}_1$ by a full rank matrix, the rank of the resulting matrix equals the rank of $\boldsymbol{H}_1$. Thus $\operatorname{rank} \boldsymbol{H}_1 = \operatorname{rank}\left(\boldsymbol{H}_1 \cdot \begin{pmatrix} \mathbf{1}_K & \boldsymbol{I}_K \\ 1 & \mathbf{0}_K^T \end{pmatrix}\right)$

$$= \operatorname{rank} \begin{pmatrix} -1 & -(H(x_{\alpha_0},x_\mu) - H(x_{\alpha_1},x_\mu)) & \dots & -(H(x_{\alpha_K},x_\mu) - H(x_{\alpha_{K+1}},x\mu)) \\ 0 & -(1-H(x_{\alpha_1},x_\mu)) & & 0 \\ \vdots & & -(1-H(x_{\alpha_2},x_\mu)) & \vdots \\ & & \ddots & 0 \\ 0 & \dots & 0 & -(1-H(x_{\alpha_K},x_\mu)) \end{pmatrix}.$$

This is an upper triangular matrix and therefore of full rank. Since $\boldsymbol{H}_1$ has full rank so has $\boldsymbol{H}$. Hereby it is shown that rank$\boldsymbol{M}_{\mu\alpha} = k+1$. □

Having a full-rank information matrix for a single observation guarantees the existence of a non-zero determinant of the information matrix for any design.

Now let us assume that only the correlation parameter τ is known and fixed. Then the information matrix for a single observation is given by

$$\boldsymbol{M}(x, \boldsymbol{\theta}^*) = \boldsymbol{V}\boldsymbol{D}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T\boldsymbol{D}\boldsymbol{V}^T$$

where $\boldsymbol{\theta}^* = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_K \\ \sigma \\ \beta \end{pmatrix}$, $\boldsymbol{D}$, $\boldsymbol{H}$ and $\boldsymbol{P}$ are as above and

$$\boldsymbol{V} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ x_\mu & 0 & \cdots & 0 \\ 0 & x_{\alpha_1} & \cdots & x_{\alpha_K} \end{pmatrix} \sim (K+3)\times(K+1).$$

The rank of $\boldsymbol{V}$ is obviously $K+1$ (full column rank), so following the same argumentation as above, $\boldsymbol{M}$ is of rank $K+1$. But since $\boldsymbol{M}$ now is of dimension $K+3$, it is not of full rank.

Full Information Matrix

The full information matrix is given by

$$\boldsymbol{M}(x, \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{V}\boldsymbol{D}\boldsymbol{H} \\ \boldsymbol{t}^T \end{pmatrix} \boldsymbol{P} \begin{pmatrix} \boldsymbol{H}^T\boldsymbol{D}\boldsymbol{V}^T & \boldsymbol{t} \end{pmatrix}$$

where $\boldsymbol{D}$, $\boldsymbol{H}$, $\boldsymbol{P}$ and $\boldsymbol{V}$ are as above and

$$\boldsymbol{t} = \begin{pmatrix} -\boldsymbol{t}_1 \\ \boldsymbol{t}_1 \end{pmatrix}$$

where

$$\boldsymbol{t}_1 = \begin{pmatrix} F(x_\mu)(1-F(x_\mu))(F(x_{\alpha_0})(1-F(x_{\alpha_0})) - F(x_{\alpha_1})(1-F(x_{\alpha_1}))) \\ F(x_\mu)(1-F(x_\mu))(F(x_{\alpha_1})(1-F(x_{\alpha_1})) - F(x_{\alpha_2})(1-F(x_{\alpha_2}))) \\ \vdots \\ F(x_\mu)(1-F(x_\mu))(F(x_{\alpha_K})(1-F(x_{\alpha_K})) - F(x_{\alpha_{K+1}})(1-F(x_{\alpha_{K+1}}))) \end{pmatrix}.$$

Again $\boldsymbol{M}$ is of rank $K+1$ and therefore not of full rank. This implies that a D-optimal design has to comprise at least two distinct designs points, since a one-point-design would lead to a singular information matrix.

Locally D-optimal designs can be constructed using an iterative algorithm. They will be identified and discussed for some specific settings in Section 5.4 The analytical determination of optimal designs is not feasible, even in the simpler case that some of the parameters are considered known.

5.4 Optimal Designs for some Special Cases

5.4.1 The Univariate Case

If only one endpoint - either efficacy or toxicity - is considered, the underlying model for the dose response relationship is the proportional odds model. In the case of a binary outcome it is equivalent to the common logistic model. D-optimal designs for these models are presented and discussed in Chapter 4.

5.4.2 The Bivariate Case

In this section, we will only consider D-optimality.

Let us first consider the model with two binary endpoints, i.e. $K = 1$. Consider the following setting with standardized parameters. Without loss of generality,

let $\sigma = 1$ and $\mu = 0$ (cf. Ford et al. (1992)). Then the D-optimal design depends on the ratio of σ and β and the difference of μ and α, as well as on τ. The D-optimal design points with the respective optimal weights for $0 \leq \alpha \leq 15$ are shown in Figures 5.3, 5.4 and B.1 to B.4 for $\beta = 1, \beta = 2$ and $\beta = 0.5$. The cases of independence ($\tau = 0$) and strong positive correlation ($\tau = 0.8$) are considered.

For the case of independence, $\tau = 0$ is considered fixed and known, and thus the designs are optimal with respect to the parameter vector $\boldsymbol{\theta}^*$, whereas for $\tau = 0.8$ the whole parameter vector $\boldsymbol{\theta}$ is considered. The designs presented are numerical approximations to the optimal designs, where the G-efficiency and, hence, D-efficiency is larger than 0.999. Thus they are very close to the true D-optimal designs.

There are some facts about these designs worth mentioning. The number of design points varies from 2 to 4 for $\tau = 0$ and from 3 to 4 for $\tau = 0.8$. It increases with increasing α and thus increasing distance between α and μ. In all considered cases, for large α the design converges to a four point design, where the difference between the design points and μ and α, respectively, converges to constants. For $\tau = 0$, the weights converge to 0.25, whereas for $\tau = 0.8$, the weights for the different design points vary.

It stands out that the D-optimal designs for $\beta = 0.5$ can be derived from the D-optimal designs for $\beta = 2$. Taking the design points of the D-optimal design for $\beta = 2$ at 2α, mirroring them at the axis $(\alpha, \frac{1}{2}\alpha)$ and dividing them by 2 yields the design points for the D-optimal design for $\beta = \frac{1}{2}$ at α. The optimal weights then are given by the optimal weights for $\beta = 2$ at 2α.

Now we will present the locally D-optimal designs for certain parameter constellations in the case of one binary endpoint and one categorical endpoint with 4 categories and thus $K = 3$. As above we will consider a standardized model where $\mu = 0$ and $\sigma = 1$. Analogously to the cases presented above, graphics are displayed for $\beta = 1$ and $\beta = 2$ and for $\tau = 0$ and $\tau = 0.8$. The value for α_1 is fixed to $0, 1$ or -1, respectively, α_2 is varied from α_1 to 10 and α_3 is given by $2\alpha_2 - \alpha_1$ to get equidistant categories. The D-optimal designs for these parameter constellations are given in Figures 5.5, 5.6 and B.5 to B.14.

For these cases, we observe a similar structure in the designs as in the case of 2×2 categories. The number of design points varies from 2 to 4 for $\tau = 0$ and from 2 to 5 for $\tau = 0.8$. It increases as the parameters α_j increase. It stands out that two of the four or five design points coincide with α_2 and α_3 for large values of these parameters, while the other design points spread out around μ and α_1. The weights corresponding to the design points equal to α_2 and α_3 converge to equal values, while the other weights can differ.

The results presented above show that for this model with a reasonable number of categories, we can derive locally optimal designs with few design points, that are thus applicable in practise.

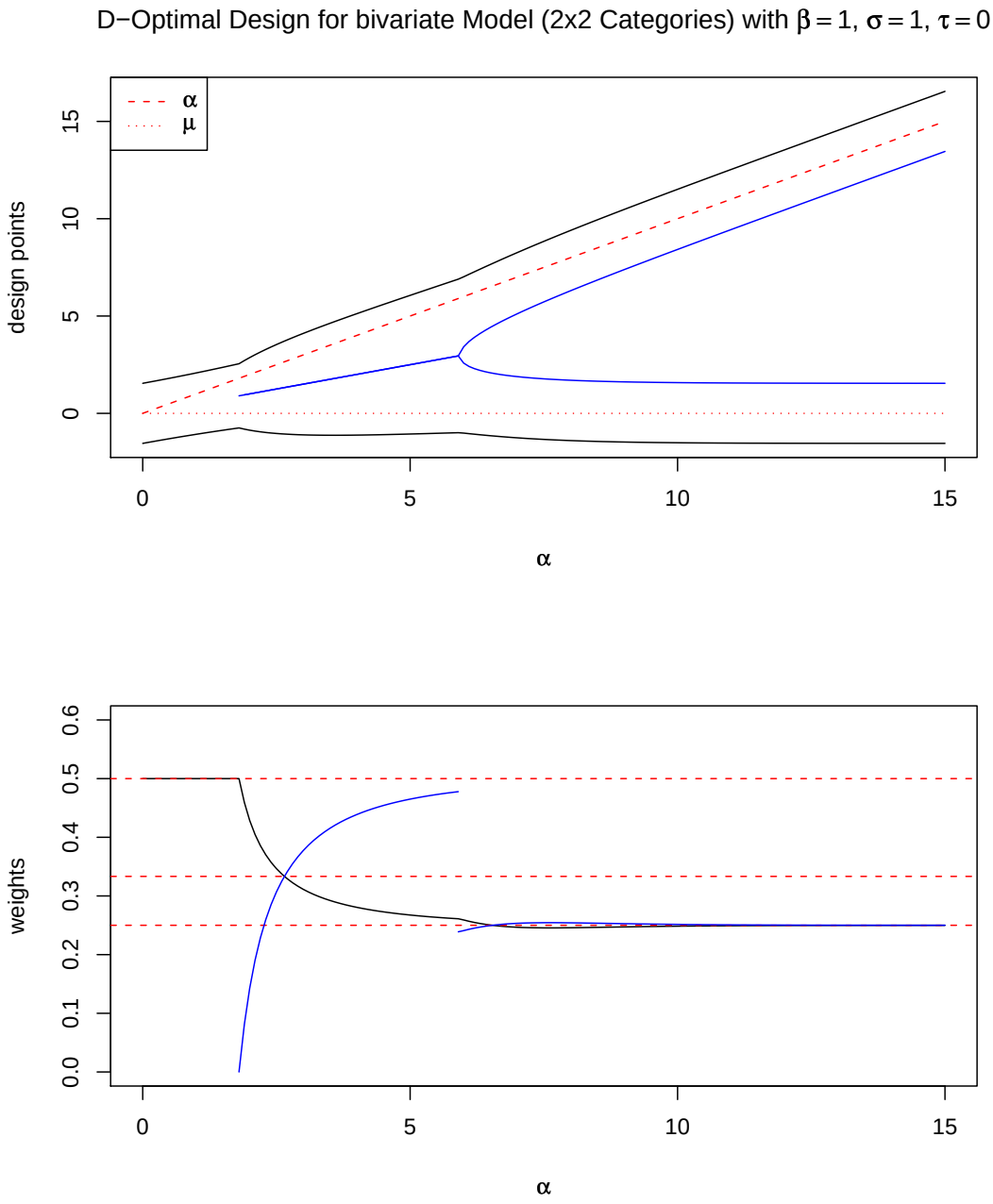


Figure 5.3: D-optimal design for the bivariate model with $\beta = 1, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

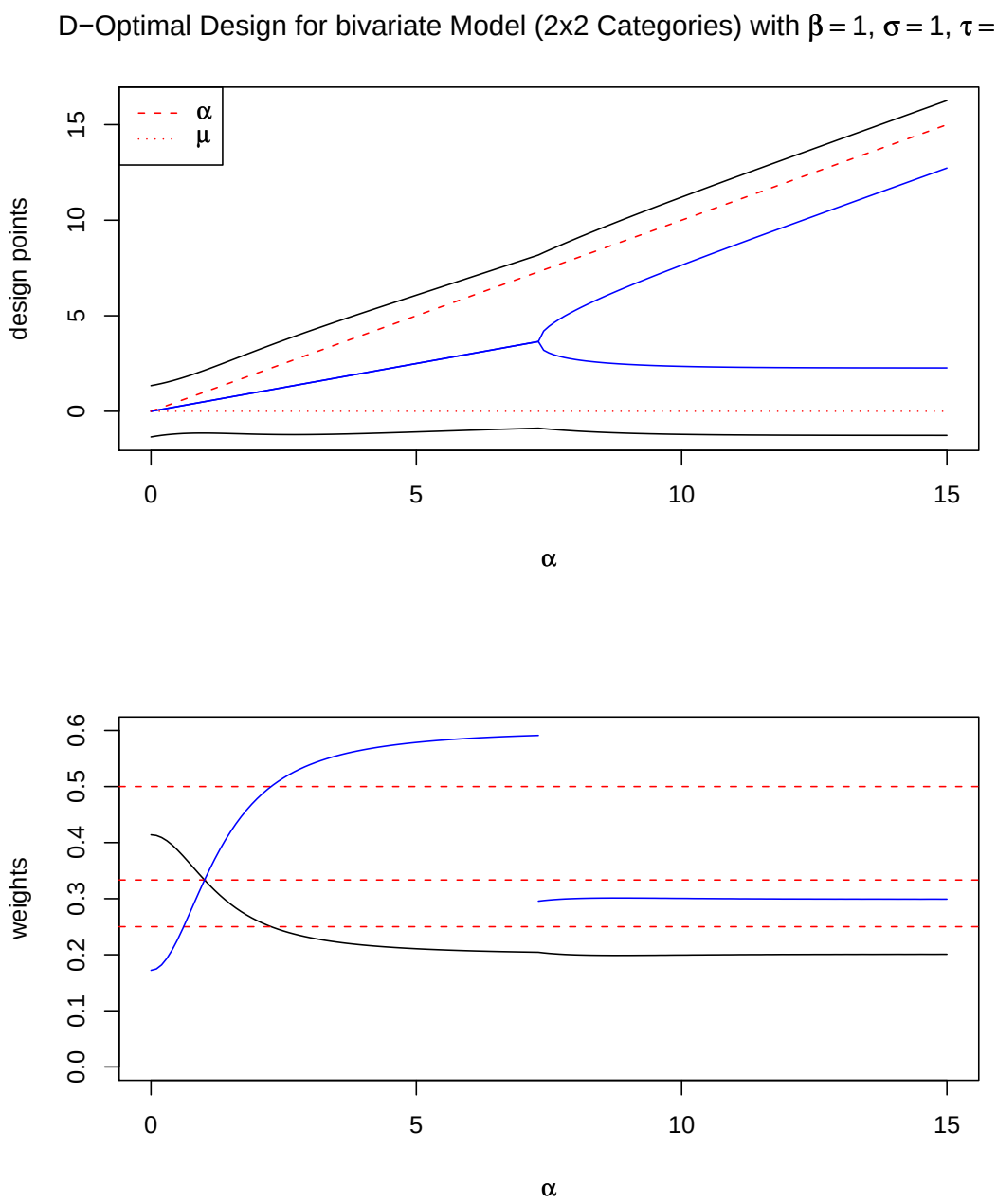


Figure 5.4: D-optimal design for the bivariate model with $\beta = 1, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

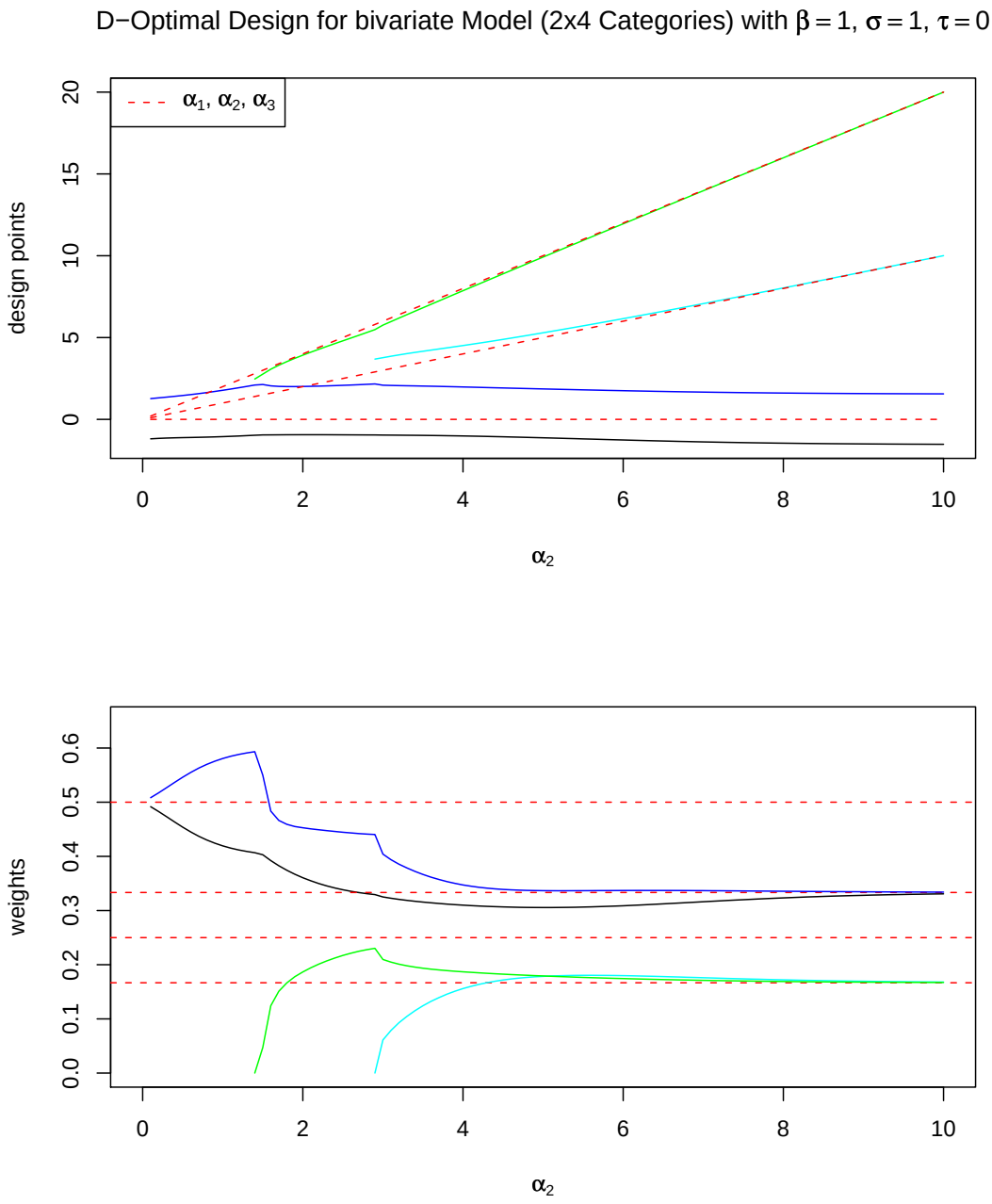


Figure 5.5: D-optimal design for the bivariate model with $\alpha_1 = 0, \beta = 1, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

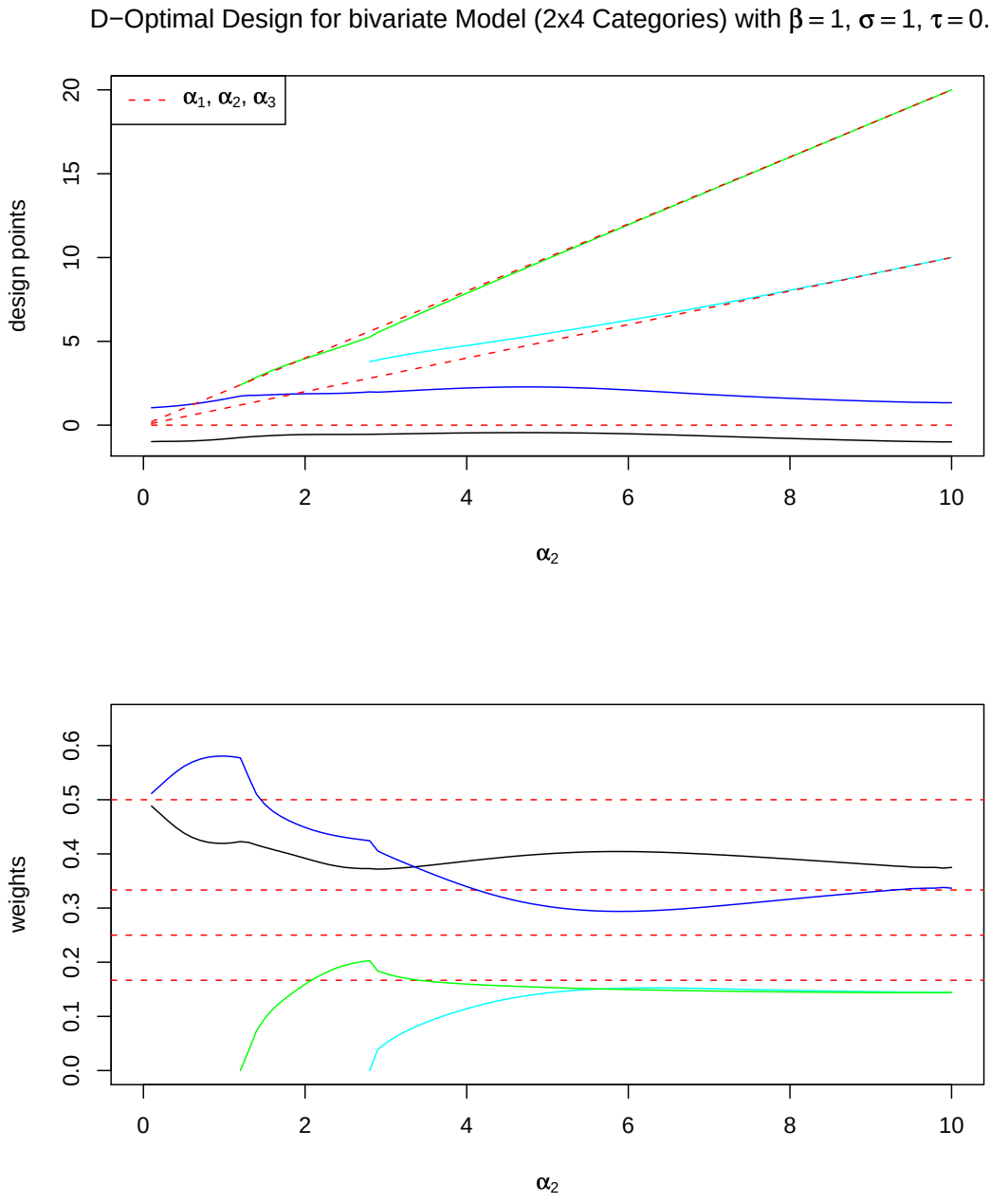


Figure 5.6: D-optimal design for the bivariate model with $\alpha_1 = 0, \beta = 2, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

6 Sequential Locally Optimal Design (SLOD)

6.1 Properties of the 3+3 Design and Problems

The 3+3 design does not depend on any model assumptions except for a monotonically increasing dose toxicity relationship. Therefore it is applicable for most realistic dose response scenarios. During the conduct of a study according to the 3+3 design, only few patients experience dose limiting toxicities and are treated with toxic doses. So the design is quite safe for the patients. On the other hand, the procedure is quite conservative, tending to underestimate the true maximally tolerated dose. Additionally, if the starting dose is very small compared to the true MTD, the number of subjects needed in the trial gets quite large. In general, the sample size depends on the course of the trial and cannot be fixed beforehand. This might have practical disadvantages since it complicates the planning of the study. These properties have been shown in various simulation studies, among others in Gerke and Siedentop (2007), and they are derived theoretically in Lin and Shih (2001).

It is not possible to make inference on the precision of the estimated MTD determined by a study conducted according to the 3+3 design without making additional model assumptions.

To avoid the disadvantages of the 3+3 design, more sophisticated methods based on optimal design theory might be appropriate.

Dragalin et al. (2006) and Lin and Shih (2001) introduce sequential approaches based on a bivariate probit model and a proportional odds model. In the following sections, we suggest a similar approach that is more flexible with respect to the model assumptions and that incorporates the 3+3 design as a start up design.

6.2 Basic Principles of SLOD

The 3+3 design is usually applied when no or very little prior information is available. In such cases, the application of locally optimal designs may not be feasible since no reliable guesses of the parameters are possible. We suggest using

the 3+3 design as a start up design to gather some information, which is then used in the construction of locally optimal designs. Thus we suggest the following approach.

Given a model appropriate for the dose response relationship (e.g. logistic, proportional odds or $E_{\max}$ model), the 3+3 design is conducted in cohorts of three patients until parameter estimation is possible in the chosen model. Then a design is determined that is locally optimal with respect to the estimated parameters, a specified optimality criterion and conditioned on the previous observations. The next cohort of patients is observed according to this design. The parameter estimation is repeated after each cohort and a new optimal design is determined. This procedure is repeated until a stopping rule is met. The stopping rule can be related to the sample size or to the precision of the estimate of the MTD. The process is displayed graphically in Figure 6.1

This procedure is quite flexible since it can be applied to different underlying models, using various optimality criteria, flexible cohort sizes and several stopping rules. The following list gives an overview of the parameters that have to be specified before the approach can be applied:

- the sequence of doses to start with
- the model
- the estimation procedure for the model parameters and the MTD
- the design region
- the optimality criterion
- the cohort size n
- the stopping rule

Given all these, the optimal design at any stage of the trial can be determined. It is given by the design that maximizes the overall information of the experiment by allocating n additional subjects to doses within the design region. The design points used so far are fixed and are denoted by the vector $\boldsymbol{x}_{obs}$. Denote the estimated parameters for the model by $\hat{\boldsymbol{\theta}}$ and the information matrix for design points $\boldsymbol{x}$ and parameters $\boldsymbol{\theta}$ by $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\theta})$. Then the conditional information matrix for the observed design points $\boldsymbol{x}_{obs}$, estimated parameters $\hat{\boldsymbol{\theta}}$ and fixed cohort size n is given by

$$\boldsymbol{M}(\boldsymbol{x}_{obs}, \hat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} \boldsymbol{M}(x_i, \hat{\boldsymbol{\theta}}), \quad x_i \in \mathcal{X}.$$

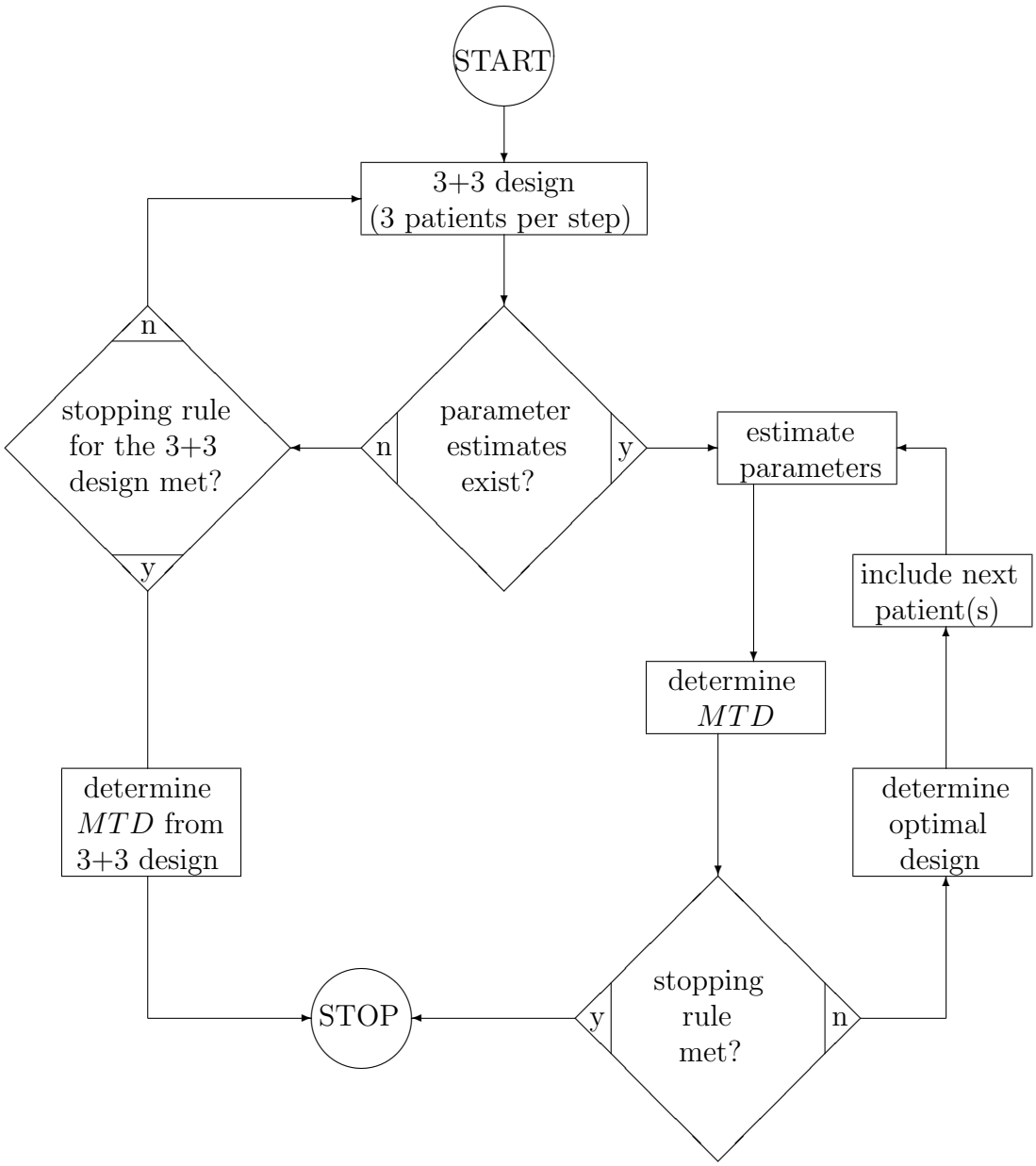


Figure 6.1: Flowchart for the Sequential Locally Optimal Design (SLOD).

Applying a design criterion to this information matrix yields the conditional optimal design. The conditional optimal design directly depends on the previous design points through their influence on the information matrix. It also depends on the outcomes at these design points through their impact on the parameter estimation.

We will now present possible parameter settings and discuss their influence on the approach. The sequence of doses to start with can be chosen in the same way as the sequence for the traditional 3+3 design is specified. The underlying model that is chosen should be the one that is considered most appropriate for the dose response relationship. We will use the common maximum likelihood method for the estimation of the model parameters.

The natural lower bound of the design region is zero. A higher value for the lower bound, e.g. the starting dose, should be chosen when the use of a placebo dose is not desired. The choice of the upper bound of the design region is not that straightforward. Choosing a value too high puts the patients at risk, while choosing a value too low unnecessarily constrains the design region. The upper bound of the design region is adaptively changed throughout the course of the trial. It should neither excessively exceed the currently estimated MTD nor the dose range used so far in the trial.

Theoretically, all doses in the interval could be used, but the design region might be restricted to a set of prespecified doses within the interval for practical reasons, e.g. when different doses cannot be supplied at short notice. The estimated MTD might as well be restricted to the set of prespecified doses. This could also be done for practical purposes as described above, or for the reason that a dose never tested should not be stated to be the MTD.

The choice of the optimality criterion depends on the primary purpose of the trial. If the interest is in only determining the MTD, a corresponding c-criterion is appropriate. If the parameter estimates for the dose response relationship are of major interest, the D-criterion is favored. The choice of the cohort size should follow practical considerations.

A very convenient stopping criterion is found by fixing the maximum sample size beforehand. This simplifies the planning of the study both in duration and cost. Nevertheless, a stopping criterion related to the precision of the estimated parameters or the MTD could be used as well.

6.3 Variance and Confidence Intervals

A major disadvantage of the 3+3 design is the lack of information on the dose response curve and on the precision of the estimate of the MTD, which is not available without any additional model assumptions. Since the method described

above is based on a parametric model, we can obtain the relevant information to determine the variance of the estimator of the MTD and thus we can calculate confidence intervals.

Assume that the dose-toxicity-relationship is described by a logistic model in the way that

$$P(DLT(x)) = \frac{\exp(\frac{x-\mu}{\sigma})}{1+\exp(\frac{x-\mu}{\sigma})}$$

with x being the dose and let $\boldsymbol{\theta} = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$. Then the MTD is the dose for which $P(DLT(MTD)) = \frac{1}{3}$, i.e. $MTD = \mu - \sigma\log(2)$. The asymptotic variance of the estimated MTD is given by

$$\mathrm{Var}(\widehat{MTD}) = \begin{pmatrix} 1 \\ -\log 2 \end{pmatrix}^T \boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta}) \begin{pmatrix} 1 \\ -\log 2 \end{pmatrix}.$$

We get an estimate for the variance by replacing $\boldsymbol{\theta}$ by a consistent estimator $\hat{\boldsymbol{\theta}}$.

Simulations show that this approximation to the variance of the estimated MTD is close to the empiric variance already for moderate sample sizes.

It can be shown by simple simulations that $\log(\widehat{MTD})$ is approximately normally distributed already for moderate sample sizes, whereas the normality assumption for $\widehat{MTD}$ is only justified for rather large sample sizes. Therefore we will construct confidence intervals for the MTD based on the log-transformation.

Let the standard error (SE) be given by

$$\mathrm{SE}(\widehat{MTD}) = \sqrt{\begin{pmatrix} 1 \\ -\log 2 \end{pmatrix}^T \boldsymbol{M}^{-1}(\xi, \hat{\boldsymbol{\theta}}) \begin{pmatrix} 1 \\ -\log 2 \end{pmatrix}}.$$

Then an approximate $(1 - \alpha)$ confidence interval for the MTD is

$$\left[\frac{\widehat{MTD}}{\exp\left(q_{1-\frac{\alpha}{2}} \frac{1}{\widehat{MTD}} \mathrm{SE}(\widehat{MTD})\right)} \, , \, \widehat{MTD} \cdot \exp\left(q_{1-\frac{\alpha}{2}} \frac{1}{\widehat{MTD}} \mathrm{SE}(\widehat{MTD})\right) \right]$$

with $q_{1-\frac{\alpha}{2}}$ being the $1-\frac{\alpha}{2}$ quantile of the standard normal distribution.

The results derived above can be transferred easily to other models with known information matrices.

For the proportional odds model with four categories, i.e.,

$$P(T(x) \geq j) = \frac{\exp(\frac{x-\alpha_j}{\beta})}{1+\exp(\frac{x-\alpha_j}{\beta})}, \; j = 1, ..., K$$

and

$$MTD = \alpha_3 - \beta \log(2)$$

we have

$$\mathrm{Var}(\widehat{MTD}) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -\log 2 \end{pmatrix}^T \boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta}) \begin{pmatrix} 1 \\ 0 \\ 0 \\ -\log 2 \end{pmatrix}.$$

With the same normality assumptions as above and

$$\mathrm{SE}(\widehat{MTD}) = \sqrt{\begin{pmatrix} 1 \\ 0 \\ 0 \\ -\log 2 \end{pmatrix}^T \boldsymbol{M}^{-1}(\xi, \hat{\boldsymbol{\theta}}) \begin{pmatrix} 1 \\ 0 \\ 0 \\ -\log 2 \end{pmatrix}},$$

an approximate $(1 - \alpha)$ confidence interval for the MTD is given by

$$\left[\frac{\widehat{MTD}}{\exp\left(q_{1-\frac{\alpha}{2}} \frac{1}{\widehat{MTD}} \mathrm{SE}(\widehat{MTD})\right)} \, , \, \widehat{MTD} \cdot \exp\left(q_{1-\frac{\alpha}{2}} \frac{1}{\widehat{MTD}} \mathrm{SE}(\widehat{MTD})\right) \right].$$

6.4 Extension to a Bivariate Setting

The method described in Section 6.2 can easily be extended to be suitable for a bivariate setting like the one presented in Chapter 5. Special attention has to be paid at the steps concerning the existence of the MLE, the estimation of the parameters and the determination of the conditional optimal design according to an appropriate design criterion.

If the main interest is in estimating all the model parameters with high precision, the D-criterion is directly applicable to the bivariate model. Another criterion is needed if the main goal of the study is to estimate the MTD and the minimum effective dose ($minED$) as precisely as possible. The MTD in the bivariate setting is defined analogously to the univariate case as the 0.33 quantile of the marginal distribution of the toxicity endpoint. The $minED$ is a certain quantile of the marginal distribution of the efficacy endpoint. Which quantile should be chosen as the $minED$ is not a statistical but medical question.

An appropriate optimality criterion for this setting is an L-criterion, i.e.

$$\min \mathrm{tr}\left(\boldsymbol{C}^T \boldsymbol{M}(\xi, \boldsymbol{\theta})^- \boldsymbol{C}\right)$$

where $\boldsymbol{C} = (C_1 \quad C_2)$ is a two-column matrix such that $\boldsymbol{C}_1^T \cdot \hat{\boldsymbol{\theta}} = \widehat{MTD}$ and $\boldsymbol{C}_2^T \cdot \hat{\boldsymbol{\theta}} = \widehat{minED}$, and $\boldsymbol{M}(\xi, \boldsymbol{\theta})^-$ is a generalized inverse of $\boldsymbol{M}(\xi, \boldsymbol{\theta})$. The information

matrix $\boldsymbol{M}(\xi, \boldsymbol{\theta})$ is not necessarily invertible, but if we confine ourselves to designs allowing for the estimation of $\boldsymbol{\theta}$, the information matrix is non-singular (see e.g. Section 5.3.1), and we can replace the generalized inverse $\boldsymbol{M}(\xi, \boldsymbol{\theta})^-$ in the above formula by $\boldsymbol{M}(\xi, \boldsymbol{\theta})^{-1}$.

The exact form of $\boldsymbol{C}$ depends on the model used and the quantile specified as $minED$. Assuming the model as described in Chapter 5 with four toxicity categories and thus $\boldsymbol{\theta} = (\mu, \alpha_1, \alpha_2, \alpha_3, \sigma, \beta, \tau)^T$, and assuming that $minED$ is the 0.33 quantile of the marginal distribution of the efficacy endpoint, we have

$$\boldsymbol{C} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ -log(2) & 0 \\ 0 & -log(2) \\ 0 & 0 \end{pmatrix}.$$

It might be necessary to observe a quite large number of subjects until parameter estimation is possible in the bivariate model. To make use of the parametric model as soon as possible, we do not want to wait until all the parameters can be estimated. A univariate model equivalent to the marginal model for toxicity can be fit to the toxicity endpoint, and as soon as the parameters in this simpler model can be estimated, we will use the SLOD as described in the previous section. As soon as all parameters in the complete bivariate model can be estimated, we switch to the bivariate optimality criterion and optimize the design according to both endpoints of interest.

7 Simulation Study

In this simulation study, the performance of the traditional 3+3 design, SLOD as presented in Chapter 6 and a Bayesian approach was compared. The goal was to find the MTD in a dose-escalation-study, when no prior information is available. Six different dose response scenarios were considered. Four of them are based on a two parameter logistic dose response relationship, while the last two scenarios assume an $E_{\max}$ model for the true dose response relationship. The scenarios are described in more detail in section 7.2, while the specific designs used are presented in Section 7.1. In Section 7.3, the results are displayed and conclusions are drawn.

7.1 Designs

The 3+3 design was used as described in Section 2.3. The sequence of doses used here is given by $d = (0.6, 1.2, 2.0, 3.0, 4.0, 5.3, 7.0, 9.3, 12.4, 16.5, 22.0, 29.4)$ where all doses are given in mg. SLOD was investigated using different settings, all with the dose sequence given above. We investigated SLOD based on a logistic model, a four category proportional odds model and a bivariate model (cf. Chapter 5) with 2×2 categories and independent endpoints ($\tau = 0$). The D-criterion and a c-criterion where $\boldsymbol{c}^T\boldsymbol{\theta} = MTD$ were used. The cohort size n was one or two. The parameters are estimated using the maximum likelihood method, and the MTD was chosen to be the largest dose out of d with estimated probability for DLT less than $\frac{1}{3}$. The lower bound of the design region was always 0, whereas the upper bound was either one dose step above the current estimated MTD (design region 1), or one dose step above the maximum dose used so far in the trial (design region 2). The upper bound of the design region though never exceeded the range of the prespecified doses d. As the stopping rule, the maximum sample size in the trial was fixed to the median sample size from the simulated 3+3 design in the corresponding scenario, to achieve comparable sample sizes.

The Bayesian approach used for the simulation study is implemented in the software tool Bayesian ADEPT (Assisted Decision Making in Early Phase Trials). This approach assumes a logistic model as the true dose reponse relationship. More details on this approach can be found in Whitehead and Williamson (1998) and Zhou and Whitehead (2002).

Scenario	μ	σ	MTD dose in mg	MTD dose step
I	30.00	7.67	22.0	11
II	23.80	8.30	16.5	10
III	50.00	14.43	29.4	12
IV	11.00	4.33	7.0	7
	ED_{50}	λ		
V	50.00	0.70	16.5	10
VI	28.00	1.60	16.5	10

Table 7.1: Parameters and MTDs for the different dose response scenarios.

For ease of comparison, the same dose sequence and cohort sizes were used as for SLOD. The default stopping rule was used, that is the trial was stopped as soon as the ratio of the upper limit of the 95% credibility interval and the corresponding lower limit was less than 5, or the maximum sample size of 60 was reached. We needed to specify the 0.25 and 0.5 quantile of the logistic dose-response curve as the prior information. The lowest and highest dose out of the specified sequence were chosen for these quantiles, respectively. The minimum possible weight was given to this prior information, that is prior information worth one observation. Two different gain functions, the variance and patient gain function were investigated.

7.2 Dose Response Scenarios

The four scenarios based on the logistic model were chosen to cover a wide range of possible dose response relationships, where the true model coincides with the model chosen for analysis. To investigate the performance of the proposed designs when the true model is different from the model chosen for analysis, the two scenarios based on the $E_{\max}$ model were considered. The scenarios were based on the logistic and the $E_{\max}$ model, since those were the models considered most appropriate for dose response relationships.

The scenarios are numbered from I to VI. Scenarios I and II are based on two different models fit to real data, whereas Scenarios III and IV represent a 'safe' and a 'toxic' scenario, respectively. Scenarios V and VI were chosen to add two dose response relationships where an approximation by a logistic model is not reasonable.

The parameters for the underlying logistic functions and the $E_{\max}$ models as well as the resulting MTDs are given in Table 7.1. Table 7.2 shows the doses used and the corresponding true probabilities of dose limiting toxicities for each of the scenarios.

Dose (in mg)	Scenario I	II	III	IV	V	VI
0.6	0.0212	0.0576	0.0316	0.0830	0.0433	0.0021
1.2	0.0229	0.0616	0.0329	0.0942	0.0684	0.0064
2.0	0.0253	0.0675	0.0347	0.1112	0.0951	0.0145
3.0	0.0287	0.0754	0.0371	0.1362	0.1225	0.0273
4.0	0.0326	0.0843	0.0396	0.1657	0.1458	0.0426
5.3	0.0384	0.0972	0.0432	0.2114	0.1721	0.0652
7.0	0.0475	0.1167	0.0483	0.2842	0.2016	0.0981
9.3	0.0630	0.1484	0.0562	0.4031	0.2355	0.1464
12.4	0.0916	0.2021	0.0688	0.5801	0.2737	0.2136
16.5	0.1473	0.2940	0.0895	0.7820	0.3154	0.3009
22.0	0.2606	0.4460	0.1256	0.9269	0.3602	0.4047
29.4	0.4805	0.6626	0.1935	0.9859	0.4081	0.5195

Table 7.2: Probabilities of dose limiting toxicity for each dose in the different scenarios.

Scenario	μ_1	μ_2
I	14.00	22.00
II	8.84	10.97
III	26.00	38.00
IV	3.00	7.00
	ED_{50}^1	ED_{50}^2
V	15.00	30.00
VI	12.00	18.00

Table 7.3: Additional parameters for the proportional odds model.

To apply SLOD based on the four category proportional odds model or the bivariate model, some more specifications are necessary.

In the part of the simulation based on the proportional odds model, the dose limiting toxicities were defined as toxicities of grade 3 or higher. Additionally all toxicities of these grades were summarized in one category, namely grade 3.

To simulate the observations in Scenario I to IV, we used proportional odds models for four categories. The slope parameter σ and the parameter μ_3, corresponding to the dose where the probability for a grade 3 toxicities is 0.5, are the same as σ and μ for the logistic model in the corresponding scenario. The additional parameters μ_1 and μ_2, corresponding to the doses where the probability of the respective lower grade toxicity is 0.5, are given in Table 7.3.

Scenario	α	β
I	20	5.0
II	18	7.0
III	20	5.0
IV	8	2.5
	ED_{50}	λ
V	40	0.5
VI	24	1.0

Table 7.4: Additional parameters for the dose efficacy relationship in the bivariate model.

For Scenarios V and VI, the concept of the proportional odds model was transferred to the $E_{\max}$ model. For each of the cumulative toxicity categories, an $E_{\max}$ model was used as the true dose toxicity relationship. The parameter λ is the same for all toxicity categories and coincides with the one used above. The parameter ED_{50}^{j} is specified for each toxicity category $j = 1, 2, 3$ with ED_{50}^{3} being identical to ED_{50} used above. The additional values for ED_{50}^{1} and ED_{50}^{2} can be seen in Table 7.3.

In the part of the simulation using the bivariate model, we used the same dose toxicity relationship as above, given by a logistic or $E_{\max}$ model respectively, and the parameters in Table 7.1. Additionally, we specified a dose efficacy relationship, following a logistic model for Scenarios I to IV and an $E_{\max}$ model for Scenarios V and VI. The parameters for these dose efficacy relationships can be found in Table 7.4. Besides the MTD, we also estimated the minimum effective dose ($minED$) as the dose x for which $P(E(x) = 1) = \frac{1}{3}$. If x is less than zero, it is set to zero, and if x exceeds the range of the prespecified doses, i.e. $x > 29$, no estimate for the minimum effective dose is available.

7.3 Results

We will now discuss the results of the simulation study. For each scenario and each setting, different characteristics are of interest. We will consider the following characteristics to quantify the precision of the estimates:

- the percentage of correctly estimated MTDs
- the distribution of the estimated MTDs
- the estimated mean squared error of the estimated MTD
- the median width of a confidence interval for the MTD.

Other quantities of interest are

- the average sample size
- the average number of observed DLTs
- the average number of patients treated with toxic doses (doses above the true MTD)
- the number of cases in which the chosen method fails to give an estimate of the MTD.

The results for each setting and each scenario are displayed in two different tables, the first one giving the percentage of each dose being estimated as the MTD, and the second table displaying the other characteristics mentioned above. The width of the confidence interval is given by the ratio of the upper and the lower limit of the approximate 95% confidence interval as derived in Chapter 6, and the median is displayed in the tables. It is only displayed for SLOD, since this confidence interval calculation is not necessarily reasonable for the other methods.

We will use the 3+3 design as the reference to which we compare the other methods.

First we will discuss the results for Scenario I. Using the 3+3 design the MTD was estimated correctly in less than 30% of the simulation runs (cf. Table 7.5). In most cases the MTD was underestimated, whereas it was overestimated in less than 8%. This systemic underestimation and the skew distribution of the estimated MTDs can also be seen in Figure 7.1. These observations are in line with the behavior of the 3+3 design described in the literature (cf. Gerke and Siedentop (2007)). In some cases (0.5%), the 3+3 design failed to give an estimate of the MTD. This occurred when several DLTs were observed on the lowest dose step. The average sample size was around 38 with an average of less than 3.5 DLTs per trial. Only few patients, on average 1.6, were treated with toxic doses. The mean squared error of the estimated MTD is 73 (cf. Table 7.9).

These observations on the distribution of the estimate of the MTD, the observed number of DLTs and the number of patients treated with toxic doses is in line with the properties of the 3+3 design described in the literature, e.g. in Gerke and Siedentop (2007).

Using the Bayesian method (cf. Tables 7.5 and 7.9), we observe a different behavior. The Bayesian approach always gave an estimate of the MTD, and this estimate was the correct MTD in 45% to 64% of the simulation runs, depending on the gain function and cohort size. The estimate was less biased than with the 3+3 design (cf. Figure 7.1), especially when the patient gain function was used, and had a considerably lower MSE with values between 16 and 25. This

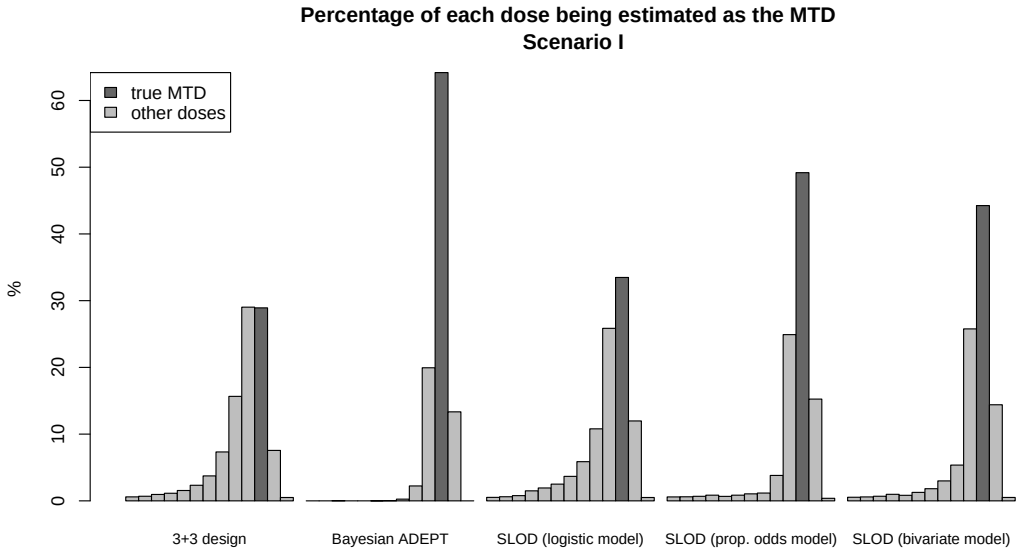


Figure 7.1: Percentage of each dose being estimated as the MTD for the different methods in Scenario I.

increase in precision comes at a certain cost: for average sample sizes a bit lower (between 28 and 34), the average number of observed toxicities (8 to 10) and the average number of patients treated at toxic doses (around 14) was considerably increased. Treating almost half of the patients with toxic doses and observing toxicities in up to one third of the patients is not tolerable in practice.

When we look at the results for SLOD (cf. Tables 7.6 and 7.10), it stands out that they differ considerably depending on the model the approach is based on. The optimality criterion though only has a minor influence on the results. The number of cases in which no MTD could be determined is around 0.5 % for most settings and even a bit lower for the method based on the proportional odds model when the cohort size is one. Using SLOD based on the logistic model, it becomes obvious that the percentage of correctly estimated MTDs is slightly higher than with the 3+3 design if design region 1 is used, and they are considerably lower for design region 2. These facts transfer directly to the magnitude of the MSE. It stands out that the width of the confidence interval is almost constant across the different settings (around 1.95), so we observe a noticeable difference in the bias of the estimates. The mean number of observed DLTs and the number of patients treated with toxic doses vary only slightly with the settings and all values are smaller than 4. Thus we only observe a minor increase in the number of DLTs compared to the 3+3 design and a moderate increase in the number of patients treated with toxic doses. Although we have a slightly increase in the risk for the patients, we do not have a noteworthy increase in the precision of the

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
0.6	0.60	0.00	0.00	0.00	0.00
1.2	0.69	0.00	0.00	0.00	0.00
2.0	0.96	0.50	0.36	0.02	0.21
3.0	1.14	0.00	0.00	0.00	0.00
4.0	1.55	0.00	0.00	0.00	0.00
5.3	2.33	0.01	0.01	0.00	0.00
7.0	3.74	0.11	0.11	0.03	0.06
9.3	7.33	0.85	1.08	0.26	0.45
12.4	15.66	6.96	7.06	2.23	2.47
16.5	29.03	40.31	38.74	19.94	22.60
22.0	**28.92**	**45.64**	**45.95**	**64.18**	**57.66**
29.4	7.56	5.62	6.69	13.34	16.5
none	0.50	0.00	0.00	0.00	0.00

Table 7.5: Percentage of each dose being estimated as the MTD in **Scenario I** for the **3+3 design** and different settings in **Bayesian ADEPT**; 100 000 simulation runs.

estimates.

Applying SLOD based on the proportional odds model, the results differ (cf. Tables 7.7 and 7.11). The influence of the different design regions almost vanishes. The percentage of correctly estimated MTDs increases to up to 50%, which directly leads to a strong reduction of the MSE by around 40% compared to the 3+3 design. The widths of the confidence intervals are comparable to the ones above, so the reduction in the MSE is owed to a reduced bias. This gain in precision is associated with a moderately increased risk for the patients of experiencing a DLT (average of 5 to 6 patients per trial) and being treated with a toxic dose (average of 5 to 7 patients per trial). Nevertheless, these numbers are still considerably smaller than those observed with the Bayesian approach.

Finally we take a look at SLOD based on a bivariate model, where not only information on toxicity but also on efficacy is observed (cf. Tables 7.8 and 7.12). As with the approach based on the logistic model, the design region has a noticeable influence on the results, where design region 1 performs better. For design region 1, the percentage of correctly estimated MTDs is considerably increased compared to the 3+3 design, but it is still lower than with the approach based on the proportional odds model. The number of observed DLTs and patients treated with toxic doses are in between those of the 3+3 design and the approach based on the proportional odds model. The widths of the confidence intervals

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	0.52	0.67	0.53	0.52	0.55	0.68	0.52	0.52
1.2	0.62	0.62	0.63	0.65	0.62	0.61	0.63	0.64
2.0	0.78	0.81	0.76	0.84	0.77	0.81	0.77	0.86
3.0	1.50	1.51	1.46	1.67	1.50	1.55	1.46	1.62
4.0	1.93	1.78	1.90	2.02	1.98	1.73	1.87	2.03
5.3	2.51	3.24	2.85	4.73	2.51	3.24	2.84	4.91
7.0	3.67	3.53	6.52	9.04	3.70	3.52	6.50	9.04
9.3	5.87	5.59	15.56	15.38	6.05	5.80	15.62	15.22
12.4	10.79	11.07	21.34	19.76	11.04	11.53	21.34	19.73
16.5	25.85	26.08	24.67	23.93	26.21	26.33	24.66	23.96
22.0	**33.48**	**30.38**	**18.09**	**15.64**	**32.54**	**29.28**	**18.11**	**15.70**
29.4	11.98	14.24	5.18	5.33	12.02	14.45	5.18	5.29
none	0.50	0.49	0.49	0.49	0.52	0.46	0.49	0.49

Table 7.6: Percentage of each dose being estimated as the MTD in **Scenario I** for **SLOD based on the logistic model** with different optimality criteria, design regions, and cohort sizes n; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	0.59	0.73	0.59	0.46	0.59	0.56	0.59	0.65
1.2	0.61	0.66	0.62	0.78	0.61	0.75	0.62	0.66
2.0	0.68	0.81	0.67	0.71	0.67	0.69	0.67	0.65
3.0	0.85	0.84	0.85	0.76	0.85	0.82	0.85	0.86
4.0	0.67	0.77	0.69	0.72	0.68	0.73	0.69	0.83
5.3	0.85	0.83	0.84	0.72	0.85	0.83	0.84	0.79
7.0	1.05	0.90	1.02	0.95	1.04	1.04	1.01	0.93
9.3	1.17	1.25	1.36	1.44	1.16	1.50	1.35	1.60
12.4	3.81	4.36	5.06	4.34	3.49	3.52	5.01	4.37
16.5	24.91	24.62	26.10	25.27	23.74	23.18	25.64	24.55
22.0	**49.18**	**45.74**	**47.72**	**48.44**	**50.53**	**48.24**	**48.26**	**48.94**
29.4	15.25	17.95	14.10	14.96	15.41	17.68	14.09	14.72
none	0.38	0.54	0.38	0.45	0.38	0.46	0.38	0.45

Table 7.7: Percentage of each dose being estimated as the MTD in **Scenario I** for **SLOD based on the 4 category proportional odds model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	0.54	0.56	0.53	0.48	0.54	0.60	0.53	0.44
1.2	0.59	0.62	0.58	0.60	0.58	0.70	0.58	0.61
2.0	0.69	0.81	0.72	0.87	0.69	0.58	0.72	0.69
3.0	0.98	0.86	0.98	1.01	0.98	0.86	0.98	0.94
4.0	0.83	0.86	0.88	0.83	0.83	1.07	0.86	1.03
5.3	1.27	1.67	1.52	2.32	1.27	1.68	1.57	2.49
7.0	1.82	1.85	3.11	5.29	1.82	1.82	3.84	6.57
9.3	2.99	2.61	10.84	12.25	3.00	2.81	12.22	13.26
12.4	5.36	6.09	20.42	18.86	5.59	6.26	20.11	18.71
16.5	25.77	25.01	27.30	26.91	28.04	26.44	26.69	26.53
22.0	**44.25**	**42.70**	**24.83**	**23.16**	**42.06**	**40.43**	**24.20**	**21.53**
29.4	14.40	15.83	7.78	6.87	14.09	16.24	7.19	6.50
none	0.51	0.53	0.51	0.55	0.51	0.51	0.51	0.70

Table 7.8: Percentage of each dose being estimated as the MTD in **Scenario I** for **SLOD based on the bivariate model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

are slightly higher than with the aforementioned approaches. On the other hand, the approach based on the bivariate model allows for additionally estimating the minimum effective dose with similar precision as the MTD.

The results for the other scenarios are displayed analogously in the tables in Appendix C.

For Scenario II, we observe the same properties when comparing the different approaches, only the exact numbers differ.

In Scenario III it stands out that the average sample size for the Bayesian approach is strongly increased, which implies that the stopping rule related to the credibility interval was rarely met and the variance of the estimates thus is considerably larger than in Scenarios I and II. This is in line with the wider confidence intervals for SLOD. The other properties of the different settings are according to those observed in the previous scenarios.

For Scenario IV, the observations related to the percentage of correctly estimated MTDs, average number of DLTs and of patients treated with toxic doses and the width of the confidence intervals are in line with the behavior in the previous scenarios. Considering the MSE, though, we observe that the settings with the highest percentage of correctly estimated MTDs no longer have the lowest MSE. Doubling the percentage of correctly estimated MTDs, as observed in one setting of SLOD based on the proportional odds model, does not decrease

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
$\overline{N}$	38.43	28.29	29.84	34.09	33.71
$\overline{N_{DLT}}$	3.44	8.06	8.48	10.20	9.80
$\overline{N_{>MTD}}$	1.61	14.16	14.71	13.87	14.01
MSE	73.05	25.22	25.20	15.89	19.84

Table 7.9: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), and mean squared error (MSE) in **Scenario I** for the **3+3 design** and different settings in **Bayesian ADEPT**.

	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	35.35	36.06	35.09	35.57	35.25	36.01	35.09	35.57
$\overline{N_{DLT}}$	3.81	3.38	3.75	3.87	3.69	3.31	3.76	3.88
$\overline{N_{>MTD}}$	3.86	2.52	3.36	3.70	3.51	2.30	3.37	3.70
MSE	69.16	72.31	97.71	108.28	70.12	73.30	97.64	108.35
CI	1.95	1.97	1.94	1.94	1.94	1.94	1.94	1.94

Table 7.10: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of the 95% confidence interval for the MTD (CI) in **Scenario I** for different settings of **SLOD based on the logistic model**.

	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	37.94	38.32	37.94	38.42	37.94	38.37	37.94	38.38
$\overline{N_{DLT}}$	5.61	5.21	5.53	5.64	5.81	5.34	5.63	5.74
$\overline{N_{>MTD}}$	6.18	4.92	5.40	5.63	6.90	5.42	5.75	5.99
MSE	39.39	42.77	40.55	39.80	38.78	41.02	40.32	40.72
CI	1.96	2.01	1.95	1.94	1.96	1.96	1.95	1.95

Table 7.11: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of the 95% confidence interval for the MTD (CI) in **Scenario I** for different settings of **SLOD based on the 4 category proportional odds model**.

	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	37.97	38.43	37.97	38.39	37.97	38.41	37.97	38.35
$\overline{N_{DLT}}$	4.81	4.24	4.55	4.58	4.68	4.11	4.59	4.65
$\overline{N_{>MTD}}$	5.42	3.81	4.12	4.31	4.87	3.46	4.23	4.55
MSE MTD	47.25	49.43	74.55	82.39	47.95	50.81	77.72	86.61
CI MTD	2.03	2.13	1.91	1.92	1.98	1.98	1.90	1.90
MSE minED	8.83	8.69	8.95	9.08	8.79	8.65	8.93	9.09
CI minED	1.87	1.85	1.85	1.86	1.86	1.86	1.86	1.86

Table 7.12: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error for the MTD (MSE MTD) and minimum effective dose (MSE minED), and median width of the 95% confidence intervals for the MTD (CI MTD) and the minimum effective dose (CI minED) in **Scenario I** for different settings of **SLOD based on the bivariate model**.

the MSE, compared to the 3+3 design. The cause for this is related to the skew distribution of the estimates combined with non-equidistant doses.

In Scenario V, it stands out that the percentage of correctly estimated MTDs with the 3+3 design is dramatically low. The behavior of the other methods in comparison with the 3+3 design does not differ substantially from those in the other scenarios. The results observed for Scenario VI are quite similar to those of Scenario II.

This simulation study shows that methods other than the 3+3 design have the potential to perform better. The Bayesian approach is best in estimating the correct MTD with lowest MSE across all scenarios, but it is not considered feasible in practice due to the strongly increased number of patients experiencing DLTs or being treated at toxic doses. This approach might become more useful in practice if the number of patients treated with excessive doses would be reduced by some sort of safety constraint. SLOD only moderately increases these numbers so that they still might be tolerable and - depending on the exact setting - performs at least slightly, but often considerably better than the 3+3 design regarding the percentage of correctly estimated MTDs and the precision and bias of the estimates. The more complex proportional odds model generally performs better than the simple logistic model, which was expected due to the more informative observations. When SLOD is used with the bivariate model, it tends to perform at least as well for the original endpoint as with the simple univariate model. The performance is improved when the MTD and the $minED$ are close to each other, resulting in similar parameters for both endpoints and thus in a better, more informative design. The bivariate model has the additional advantage of being able to estimate the parameters related to the second endpoint with considerable precision.

8 Discussion and Outlook

In this thesis, we have applied optimal design theory to the area of clinical studies, in particular to dose finding studies. We have introduced commonly used models from this field of application (logistic, the proportional odds and $E_{\max}$ model). For the logistic and the proportional odds model, we have derived conditions for the existence of the maximum likelihood estimator and have presented locally optimal designs for these models.

We have developed a new bivariate model that allows for simultaneous modelling of a binary efficacy endpoint and a categorical toxicity endpoint. We have applied common optimal design theory to derive the approximate Fisher information matrix and locally optimal designs for this model. For certain parameter constellations we have presented locally D-optimal designs and we have shown how the D-optimal designs depend on the parameters. It stands out that the D-optimal designs comprise only few design points, varying between two and five depending on the parameter constellation. This gives the opportunity to find exact designs that come close to the optimal continuous designs and are realizable in practice.

In practical application we often deal with situations where reliable prior information on the parameters is not available and thus a sequential procedure is favorable. Thus, to make the results described above applicable, we have introduced a sequential approach combining the standard 3+3 design with locally optimal designs, which we call the 'Sequential locally optimal design' ('SLOD'). There we use the 3+3 design as a start-up design until it is possible to fit the desired parametric model. This model then is used to determine a locally optimal design, which is applied to the next cohort of patients. After each cohort, the parameter estimates are updated and the locally optimal design is adjusted.

To investigate the properties of the suggested method, we have conducted a simulation study comparing SLOD to the standard 3+3 design and a Bayesian approach. In this simulation study, we investigated SLOD based on three different models, namely the common logistic model, the proportional odds model with four categories and the bivariate model introduced in Chapter 5.

The results of the simulation study affirm the properties of the 3+3 design described in the literature (Gerke and Siedentop (2007), Lin and Shih (2001)). Additionally, they show that the Bayesian approach as investigated is not applicable in practice due to the excessive risk for the patients. It can be seen

that SLOD has the potential to perform better than the traditional 3+3 design in finding the correct maximum tolerated dose (MTD) and giving a less biased estimate of the MTD. With respect to these criteria, SLOD is superior to the 3+3 design especially when it is applied with the more complex proportional odds model or in a bivariate setting. As opposed to the traditional 3+3 design, SLOD allows for meaningful conclusions on the precision of the estimates and the dose response relationship. Using the bivariate model, we get valuable additional information on the dose efficacy relationship without noteworthy loss in information with respect to the dose toxicity relationship. However, as compared to the standard 3+3 design, SLOD is not only considerably more complex to apply, it also slightly to moderately increases the risk for the patients of suffering a dose limiting toxicity (DLT) or being treated at toxic doses. Thus we have to consider carefully the trade off between the gain in precision and the increased patient risk.

Even though the simulation study considered several dose response scenarios and a variety of settings, the performance of the suggested method was only investigated for a limited number of cases. Since this method is very flexible, it could still be investigated if e.g. the risk for the patients can be reduced while retaining the precision of the estimates by choosing different design regions or by applying a different optimality criterion. It might also be worth investigating different models for the analysis of the dose toxicity relationship, e.g. the $E_{\max}$-model. Nevertheless, there will always be a trade off between the precision of the estimates and the risk for the patients, but this can still be optimized. Simulations as conducted within this work are essential for optimizing this method and investigating its performance in realistic scenarios. Thus they provide a valuable tool to help both the statistician and clinician in choosing a possibly new method for conducting a study.

We have seen that the parametric modification of the 3+3 design is a promising alternative to the traditional 3+3 design and to Bayesian approaches. It might be worth investigating this approach further in future work, optimizing it and last but not least applying it not only in simulations but in actual dose escalation studies.

A Derivation of the Information Matrix for the Bivariate Model

The information matrix presented in Chapter 5 is based on the following calculations, that yield the individual elements. Recall that the information matrix is symmetric and thus $\boldsymbol{M}_{ij} = \boldsymbol{M}_{ji}$.

To calculate the information matrix, let us first consider the derivatives of the used functions. For notational convenience, define

$$x_\mu := \frac{x-\mu}{\sigma}, \quad x_{\alpha_j} := \frac{x-\alpha_j}{\beta} \text{ and}$$

$$H(x,y) := F(x)(1+\tau(1-F(x))(1-2F(y))).$$

Note that

$$\frac{\partial}{\partial y}G(x,y) = F(y)(1-F(y))H(x,y).$$

Then we get

$$\begin{aligned}
\frac{\partial}{\partial\mu}F(x_\mu) &= -\frac{1}{\sigma}F(x_\mu)(1-F(x_\mu)) \\
\frac{\partial}{\partial\sigma}F(x_\mu) &= -\frac{x-\mu}{\sigma^2}F(x_\mu)(1-F(x_\mu)) \\
&= x_\mu\frac{\partial}{\partial\mu}F(x_\mu) \\
\frac{\partial}{\partial\alpha_j}F(x_{\alpha_j}) &= -\frac{1}{\beta}F(x_{\alpha_j})(1-F(x_{\alpha_j})) \\
\frac{\partial}{\partial\beta}F(x_{\alpha_j}) &= -\frac{1}{\beta}x_{\alpha_j}F(x_{\alpha_j})(1-F(x_{\alpha_j})) \\
&= x_{\alpha_j}\frac{\partial}{\partial\alpha_j}F(x_{\alpha_j})
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial\mu}G(x_{\alpha_j},x_\mu) &= -\frac{1}{\sigma}F(x_\mu)(1-F(x_\mu))H(x_{\alpha_j},x_\mu)\\
\frac{\partial}{\partial\sigma}G(x_{\alpha_j},x_\mu) &= -\frac{x-\mu}{\sigma^2}F(x_\mu)(1-F(x_\mu))H(x_{\alpha_j},x_\mu)\\
&= x_\mu\frac{\partial}{\partial\mu}G(x_{\alpha_j},x_\mu)\\
\frac{\partial}{\partial\alpha_j}G(x_{\alpha_j},x_\mu) &= -\frac{1}{\beta}F(x_{\alpha_j})(1-F(x_{\alpha_j}))H(x_\mu,x_{\alpha_j})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\beta}G(x_{\alpha_j},x_\mu) &= -\frac{1}{\beta}x_{\alpha_j}F(x_{\alpha_j})(1-F(x_{\alpha_j}))H(x_\mu,x_{\alpha_j})\\
&= x_{\alpha_j}\frac{\partial}{\partial\alpha_j}G(x_{\alpha_j},x_\mu)\\
\frac{\partial}{\partial\tau}G(x_{\alpha_j},x_\mu) &= F(x_{\alpha_j})(1-F(x_{\alpha_j}))F(x_\mu)(1-F(x_\mu)).
\end{aligned}$$

Note that $H(x_{\alpha_0},x_\mu)=1$ and $H(x_{\alpha_{K+1}},x_\mu)=0$.

All other derivatives of $F(x_\mu)$, $F(x_{\alpha_j})$ and $G(x_{\alpha_j},x_\mu)$ with respect to the elements of $\boldsymbol{\theta}$ are 0.

We have

$$\begin{aligned}
p_{j0} &= F(x_{\alpha_j})-F(x_{\alpha_{j+1}})-G(x_{\alpha_j},x_\mu)+G(x_{\alpha_{j+1}},x_\mu)\\
p_{j1} &= G(x_{\alpha_j},x_\mu)-G(x_{\alpha_{j+1}},x_\mu),
\end{aligned}$$

and thus we get

$$\begin{aligned}
\frac{\partial p_{j0}}{\partial\mu} &= -\frac{\partial}{\partial\mu}(G(x_{\alpha_j},x_\mu))+\frac{\partial}{\partial\mu}G(x_{\alpha_{j+1}},x_\mu)\\
&= \frac{1}{\sigma}F(x_\mu)(1-F(x_\mu))H(x_{\alpha_j},x_\mu)-\frac{1}{\sigma}F(x_\mu)(1-F(x_\mu))H(x_{\alpha_{j+1}},x_\mu)\\
&= \frac{1}{\sigma}F(x_\mu)(1-F(x_\mu))\left[H(x_{\alpha_j},x_\mu)-H(x_{\alpha_{j+1}},x_\mu)\right]\\
\frac{\partial p_{j1}}{\partial\mu} &= -\frac{1}{\sigma}F(x_\mu)(1-F(x_\mu))\left[H(x_{\alpha_j},x_\mu)-H(x_{\alpha_{j+1}},x_\mu)\right]\\
&= -\frac{\partial p_{j0}}{\partial\mu}\\
\frac{\partial p_{k0}}{\partial\alpha_j} &= \begin{cases} -\frac{1}{\beta}F(x_{\alpha_j})(1-F(x_{\alpha_j}))\left(1-H(x_\mu,x_{\alpha_j})\right) & , \quad k=j\\ \frac{1}{\beta}F(x_{\alpha_j})(1-F(x_{\alpha_j}))\left(1-H(x_\mu,x_{\alpha_j})\right) & , \quad k+1=j\\ 0 & , \quad \text{otherwise}\end{cases}\\
\frac{\partial p_{k1}}{\partial\alpha_j} &= \begin{cases} -\frac{1}{\beta}F(x_{\alpha_j})(1-F(x_{\alpha_j}))H(x_\mu,x_{\alpha_j}) & , \quad k=j\\ \frac{1}{\beta}F(x_{\alpha_j})(1-F(x_{\alpha_j}))H(x_\mu,x_{\alpha_j}) & , \quad k+1=j\\ 0 & , \quad \text{otherwise}\end{cases}.
\end{aligned}$$

The elements of $\boldsymbol{M}$ related to the parameters μ and $\alpha_j, j = 1, ..., K$ are given by

$$\begin{aligned}
\boldsymbol{M}_{\mu\mu} &= \sum_{i,j} \frac{1}{p_{ji}} \frac{\partial p_{ji}}{\partial \mu} \frac{\partial p_{ji}}{\partial \mu} \\
&= \sum_{j=0}^{K} \left[\frac{1}{p_{j0}} \left(\frac{\partial p_{j0}}{\partial \mu} \right)^2 + \frac{1}{p_{j1}} \left(\frac{\partial p_{j1}}{\partial \mu} \right)^2 \right] \\
&= \sum_{j=0}^{K} \left(\frac{1}{p_{j0}} + \frac{1}{p_{j1}} \right) \frac{1}{\sigma^2} F(x_\mu)^2 (1 - F(x_\mu))^2 \left[H(x_{\alpha_j}, x_\mu) - H(x_{\alpha_{j+1}}, x_\mu) \right]^2 \\
\boldsymbol{M}_{\mu\alpha_j} &= \sum_{k=0}^{K} \left[\frac{1}{p_{k0}} \frac{\partial p_{k0}}{\partial \mu} \frac{\partial p_{k0}}{\partial \alpha_j} + \frac{1}{p_{k1}} \frac{\partial p_{k1}}{\partial \mu} \frac{\partial p_{k1}}{\partial \alpha_j} \right] \\
\boldsymbol{M}_{\mu\alpha_j} &= \frac{1}{\sigma} \frac{1}{\beta} F(x_\mu)(1 - F(x_\mu) F(x_{\alpha_j})(1 - F(x_{\alpha_j})) \\
&\quad \cdot \left[\left(\frac{1}{p_{(j-1)0}} (H(x_{\alpha_{j-1}}, x_\mu) - H(x_{\alpha_j}, x_\mu)) - \frac{1}{p_{j0}} (H(x_{\alpha_j}, x_\mu) - H(x_{\alpha_{j+1}}, x_\mu)) \right) \right. \\
&\quad \cdot \left(1 - H(x_\mu, x_{\alpha_j}) \right) \\
&\quad - \left(\frac{1}{p_{(j-1)1}} (H(x_{\alpha_{j-1}}, x_\mu) - H(x_{\alpha_j}, x_\mu)) - \frac{1}{p_{j1}} (H(x_{\alpha_j}, x_\mu) - H(x_{\alpha_{j+1}}, x_\mu)) \right) \\
&\quad \left. \cdot H(x_\mu, x_{\alpha_j}) \right]
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{M}_{\alpha_j\alpha_j} &= \sum_{k=0}^{K} \left[\frac{1}{p_{k0}} \left(\frac{\partial p_{k0}}{\partial \alpha_j} \right)^2 + \frac{1}{p_{k1}} \left(\frac{\partial p_{k1}}{\partial \alpha_j} \right)^2 \right] \\
&= \frac{1}{\beta^2} F(x_{\alpha_j})^2 (1 - F(x_{\alpha_j}))^2 \cdot \\
&\quad \left[\left(\frac{1}{p_{(j-1)0}} + \frac{1}{p_{j0}} \right) \left(1 - H(x_\mu, x_{\alpha_j}) \right)^2 + \left(\frac{1}{p_{(j-1)1}} + \frac{1}{p_{j1}} \right) H(x_\mu, x_{\alpha_j})^2 \right]
\end{aligned}$$

$$\boldsymbol{M}_{\alpha_j\alpha_i} = \sum_{k=0}^{K} \left[\frac{1}{p_{k0}} \frac{\partial p_{k0}}{\partial \alpha_j} \frac{\partial p_{k0}}{\partial \alpha_i} + \frac{1}{p_{k1}} \frac{\partial p_{k1}}{\partial \alpha_j} \frac{\partial p_{k1}}{\partial \alpha_i} \right] = \begin{cases} \boldsymbol{M}_{\alpha_j\alpha_j} & , \quad i = j \\ \boldsymbol{M}_{\alpha_j\alpha_{j-1}} & , \quad i = j - 1 \\ \boldsymbol{M}_{\alpha_j\alpha_{j+1}} & , \quad i = j + 1 \\ 0 & , \quad |j - i| \geq 2 \end{cases}$$

Note that $\boldsymbol{M}_{\alpha_j\alpha_{j-1}} = \boldsymbol{M}_{\alpha_{j'+1}\alpha_{j'}} = \boldsymbol{M}_{\alpha_{j'}\alpha_{j'+1}}$ with $j' = j - 1$. Thus it suffices to consider either $\boldsymbol{M}_{\alpha_j\alpha_{j-1}}$ or $\boldsymbol{M}_{\alpha_j\alpha_{j+1}}$.

$$
\begin{aligned}
\boldsymbol{M}_{\alpha_j\alpha_{j+1}} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial\alpha_j}\frac{\partial p_{k0}}{\partial\alpha_{j+1}}+\frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial\alpha_j}\frac{\partial p_{k1}}{\partial\alpha_{j+1}}\right]\\
&= \frac{1}{\beta^2}F(x_{\alpha_j})(1-F(x_{\alpha_j}))F(x_{\alpha_{j+1}})(1-F(x_{\alpha_{j+1}}))\cdot\\
&\quad\left[-\frac{1}{p_{j0}}\left(1-H(x_\mu,x_{\alpha_j})\right)\left(1-H(x_\mu,x_{\alpha_{j+1}})\right)-\frac{1}{p_{j1}}H(x_\mu,x_{\alpha_j})H(x_\mu,x_{\alpha_{j+1}})\right].
\end{aligned}
$$

Thus we get

$$
\boldsymbol{M}\left(x,\begin{pmatrix}\mu\\ \alpha\end{pmatrix}\right)=\boldsymbol{DHPH}^T\boldsymbol{D}
$$

with $\boldsymbol{D}$, $\boldsymbol{H}$ and $\boldsymbol{P}$ as defined in Chapter 5. This can be seen directly by expanding the above matrix terms.

Now we take a look at the elements related to the parameters σ and β. The required derivatives are

$$
\begin{aligned}
\frac{\partial p_{k0}}{\partial\sigma} &= -\frac{\partial}{\partial\sigma}G(x_{\alpha_k},x_\mu)+\frac{\partial}{\partial\sigma}G(x_{\alpha_{k+1}},x_\mu)\\
&= \frac{1}{\sigma}x_\mu F(x_\mu)(1-F(x_\mu))\left[H(x_{\alpha_k},x_\mu)-H(x_{\alpha_{k+1}},x_\mu)\right]\\
\frac{\partial p_{k1}}{\partial\sigma} &= \frac{\partial}{\partial\sigma}G(x_{\alpha_k},x_\mu)-\frac{\partial}{\partial\sigma}G(x_{\alpha_{k+1}},x_\mu)\\
&= -\frac{1}{\sigma}x_\mu F(x_\mu)(1-F(x_\mu))\left[H(x_{\alpha_k},x_\mu)-H(x_{\alpha_{k+1}},x_\mu)\right]
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial p_{k0}}{\partial\beta} &= \frac{\partial}{\partial\beta}\left(F(x_{\alpha_k})-F(x_{\alpha_{k+1}})-G(x_{\alpha_k},x_\mu)+G(x_{\alpha_{k+1}},x_\mu)\right)\\
&= -\frac{1}{\beta}x_{\alpha_k}F(x_{\alpha_k})(1-F(x_{\alpha_k}))\left[1-H(x_\mu,x_{\alpha_k})\right]\\
&\quad+\frac{1}{\beta}x_{\alpha_{k+1}}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\left[1-H(x_\mu,x_{\alpha_{k+1}})\right]\\
\frac{\partial p_{k1}}{\partial\beta} &= -\frac{1}{\beta}x_{\alpha_k}F(x_{\alpha_k})(1-F(x_{\alpha_k}))H(x_\mu,x_{\alpha_k})\\
&\quad+\frac{1}{\beta}x_{\alpha_{k+1}}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))H(x_\mu,x_{\alpha_{k+1}}).
\end{aligned}
$$

It can easily be seen that

$$
\begin{aligned}
\frac{\partial p_{ki}}{\partial\sigma} &= x_\mu\frac{\partial p_{ki}}{\partial\mu}\quad\text{and}\\
\frac{\partial p_{ki}}{\partial\beta} &= x_{\alpha_k}\frac{\partial p_{ki}}{\partial\alpha}+x_{\alpha_{k+1}}\frac{\partial p_{ki}}{\partial\alpha_{k+1}}.
\end{aligned}
$$

This gives

$$\begin{aligned}
\boldsymbol{M}_{\mu\sigma} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial\mu}\frac{\partial p_{k0}}{\partial\sigma}+\frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial\mu}\frac{\partial p_{k1}}{\partial\sigma}\right]\\
&= \frac{1}{\sigma^2}x_\mu F(x_\mu)^2(1-F(x_\mu))^2\sum_{k=0}^{K}\left(\frac{1}{p_{k0}}+\frac{1}{p_{k1}}\right)\left[H(x_{\alpha_k},x_\mu)-H(x_{\alpha_{k+1}},x_\mu)\right]\\
&= x_\mu\boldsymbol{M}_{\mu\mu}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{M}_{\alpha_j\sigma} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial\alpha_j}\frac{\partial p_{k0}}{\partial\sigma}+\frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial\alpha_j}\frac{\partial p_{k1}}{\partial\sigma}\right]\\
&= \frac{1}{\sigma}\frac{1}{\beta}x_\mu F(x_\mu)(1-F(x_\mu)F(x_{\alpha_j})(1-F(x_{\alpha_j}))\\
&\quad\cdot\left[\left(\frac{1}{p_{(j-1)0}}(H(x_{\alpha_{j-1}},x_\mu)-H(x_{\alpha_j},x_\mu))-\frac{1}{p_{j0}}(H(x_{\alpha_j},x_\mu)-H(x_{\alpha_{j+1}},x_\mu))\right)\right.\\
&\quad\cdot\left(1-H(x_\mu,x_{\alpha_j})\right)\\
&\quad-\left(\frac{1}{p_{(j-1)1}}(H(x_{\alpha_{j-1}},x_\mu)-H(x_{\alpha_j},x_\mu))-\frac{1}{p_{j1}}(H(x_{\alpha_j},x_\mu)-H(x_{\alpha_{j+1}},x_\mu))\right)\\
&\quad\left.\cdot H(x_\mu,x_{\alpha_j})\right]\\
&= x_\mu\boldsymbol{M}_{\mu\alpha_j}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{M}_{\sigma\sigma} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\left(\frac{\partial p_{k0}}{\partial\sigma}\right)^2+\frac{1}{p_{k1}}\left(\frac{\partial p_{k1}}{\partial\sigma}\right)^2\right]\\
&= (x_\mu)^2\frac{1}{\sigma^2}F(x_\mu)^2(1-F(x_\mu)^2\sum_{k=0}^{K}\left(\frac{1}{p_{j0}}+\frac{1}{p_{j1}}\right)\left[H(x_{\alpha_k},x_\mu)-H(x_{\alpha_{k+1}},x_\mu)\right]^2\\
&= (x_\mu)^2\boldsymbol{M}_{\mu\mu}
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{M}_{\mu\beta} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial\mu}\frac{\partial p_{k0}}{\partial\beta}+\frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial\mu}\frac{\partial p_{k1}}{\partial\beta}\right]\\
&= \frac{1}{\sigma}\frac{1}{\beta}F(x_\mu)(1-F(x_\mu))\sum_{k=0}^{K}\left[H(x_{\alpha_k},x_\mu)-H(x_{\alpha_{k+1}},x_\mu)\right]\\
&\quad\cdot\left[\frac{1}{p_{k0}}\left(-x_{\alpha_k}F(x_{\alpha_k})(1-F(x_{\alpha_k}))\left[1-H(x_\mu,x_{\alpha_k})\right]\right.\right.\\
&\quad\left.+x_{\alpha_{k+1}}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\left[1-H(x_\mu,x_{\alpha_{k+1}})\right]\right)\\
&\quad+\frac{1}{p_{k1}}\left(x_{\alpha_k}F(x_{\alpha_k})(1-F(x_{\alpha_k}))H(x_\mu,x_{\alpha_k})\right.\\
&\quad\left.\left.-x_{\alpha_{k+1}}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))H(x_\mu,x_{\alpha_{k+1}})\right)\right].
\end{aligned}
$$

By renumeration of the sum and by using the fact that the summands are zero for $k\leq 0$ and $k>K$, we can conclude

$$\boldsymbol{M}_{\mu\beta}=\sum_{k=1}^{K}x_{\alpha_k}\boldsymbol{M}_{\mu\alpha_k}.$$

Following similar argumentation we get

$$
\begin{aligned}
\boldsymbol{M}_{\alpha_j\beta} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial\alpha_j}\frac{\partial p_{k0}}{\partial\beta}+\frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial\alpha_j}\frac{\partial p_{k1}}{\partial\beta}\right]\\
&= \frac{1}{\beta^2}\sum_{k=0}^{K}F(x_{\alpha_k})(1-F(x_{\alpha_k}))\\
&\quad\left\{(1-H(x_\mu,x_{\alpha_k}))\left[-x_{\alpha_{k-1}}\frac{1}{p_{(k-1)0}}F(x_{\alpha_{k-1}})(1-F(x_{\alpha_{k-1}}))\left(1-H(x_\mu,x_{\alpha_{k-1}})\right)\right.\right.\\
&\quad+x_{\alpha_k}\left(\frac{1}{p_{(k-1)0}}+\frac{1}{p_{k0}}\right)F(x_{\alpha_k})(1-F(x_{\alpha_k}))\left(1-H(x_\mu,x_{\alpha_k})\right)\\
&\quad\left.-x_{\alpha_{k+1}}\frac{1}{p_{k0}}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\left(1-H(x_\mu,x_{\alpha_{k+1}})\right)\right]\\
&\quad+H(x_\mu,x_{\alpha_k})\left[-x_{\alpha_{k-1}}\frac{1}{p_{(k-1)1}}F(x_{\alpha_{k-1}})(1-F(x_{\alpha_{k-1}}))H(x_\mu,x_{\alpha_{k-1}})\right.\\
&\quad+x_{\alpha_k}\left(\frac{1}{p_{(k-1)1}}+\frac{1}{p_{k1}}\right)F(x_{\alpha_k})(1-F(x_{\alpha_k}))H(x_\mu,x_{\alpha_k})\\
&\quad\left.\left.-x_{\alpha_{k+1}}\frac{1}{p_{k1}}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))1-H(x_\mu,x_{\alpha_{k+1}})\right]\right\}\\
&= \sum_{k=1}^{K}x_{\alpha_k}\boldsymbol{M}_{\mu\alpha_k}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{M}_{\sigma\beta} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial\sigma}\frac{\partial p_{k0}}{\partial\beta}+\frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial\sigma}\frac{\partial p_{k1}}{\partial\beta}\right]\\
&= \frac{1}{\beta}\frac{1}{\sigma}x_\mu F(x_\mu)(1-F(x_\mu))\sum_{k=0}^{K}\left[H(x_{\alpha_k},x_\mu)-H(x_{\alpha_{k+1}},x_\mu)\right]\\
&\quad\left\{F(x_{\alpha_k})(1-F(x_{\alpha_k}))\frac{x-\alpha_k}{\beta}\left(-\frac{1}{p_{k0}}\left[1-H(x_\mu,x_{\alpha_k})\right]+\frac{1}{p_{k1}}H(x_\mu,x_{\alpha_k})\right)\right.\\
&\quad\left.+(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\frac{x-\alpha_{k+1}}{\beta}\left(-\frac{1}{p_{k0}}\left[1-H(x_\mu,x_{\alpha_{k+1}})\right]+\frac{1}{p_{k1}}H(x_\mu,x_{\alpha_{k+1}})\right)\right\}\\
&= x_\mu\boldsymbol{M}_{\mu\beta}\\
&= x_\mu\sum_{k=0}^{K}x_{\alpha_k}\boldsymbol{M}_{\mu\alpha_k}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{M}_{\beta\beta} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\left(\frac{\partial p_{k0}}{\partial\beta}\right)^2+\frac{1}{p_{k1}}\left(\frac{\partial p_{k1}}{\partial\beta}\right)^2\right]\\
&= \sum_{k=0}^{K}\left\{\frac{1}{p_{k0}}\left[-\frac{x-\alpha_k}{\beta^2}F(x_{\alpha_k})(1-F(x_{\alpha_k}))\left[1-H(x_\mu,x_{\alpha_k})\right]\right.\right.\\
&\quad\left.+\frac{x-\alpha_{k+1}}{\beta^2}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\left[1-H(x_\mu,x_{\alpha_{k+1}})\right]\right]^2\\
&\quad+\frac{1}{p_{k1}}\left[-\frac{x-\alpha_k}{\beta^2}F(x_{\alpha_k})(1-F(x_{\alpha_k}))H(x_\mu,x_{\alpha_k})\right.\\
&\quad\left.\left.+\frac{x-\alpha_{k+1}}{\beta^2}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))H(x_\mu,x_{\alpha_{k+1}})\right]^2\right\}\\
&= \sum_{j=0}^{K}x_{\alpha_j}\left[x_{\alpha_{j-1}}\boldsymbol{M}_{\alpha_{j-1}\alpha_j}+x_{\alpha_j}\boldsymbol{M}_{\alpha_j\alpha_j}+x_{\alpha_{j+1}}\boldsymbol{M}_{\alpha_{j+1}\alpha_j}\right]\\
&= \sum_{j=0}^{K}x_{\alpha_j}\sum_{k=0}^{K}x_{\alpha_k}\boldsymbol{M}_{\alpha_j\alpha_k}.
\end{aligned}
$$

From these results we get directly

$$
\boldsymbol{M}\left(x,\begin{pmatrix}\mu\\ \alpha\\ \beta\\ \sigma\end{pmatrix}\right)=\boldsymbol{VDHPH}^T\boldsymbol{DV}^T
$$

with

$$\boldsymbol{V} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ x_\mu & 0 & \cdots & 0 \\ 0 & x_{\alpha_1} & \cdots & x_{\alpha_K} \end{pmatrix} \sim (K+3) \times (K+1).$$

For the full information matrix, we still need to consider the elements of $\boldsymbol{M}$ related to τ.

$$\begin{aligned} \frac{\partial p_{k0}}{\partial \tau} &= \frac{\partial}{\partial \tau}\left(F(x_{\alpha_k}) - F(x_{\alpha_{k+1}}) - G(x_{\alpha_k}, x_\mu) + G(x_{\alpha_{k+1}}, x_\mu)\right) \\ &= F(x_\mu)(1-F(x_\mu))\left[-F(x_{\alpha_k})(1-F(x_{\alpha_k})) + F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\right] \\ \frac{\partial p_{k1}}{\partial \tau} &= F(x_\mu)(1-F(x_\mu))\left[F(x_{\alpha_k})(1-F(x_{\alpha_k})) - F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\right] \end{aligned}$$

$$\begin{aligned} \boldsymbol{M}_{\mu\tau} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial \mu}\frac{\partial p_{k0}}{\partial \tau} + \frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial \mu}\frac{\partial p_{k1}}{\partial \tau}\right] \\ &= \frac{1}{\sigma}F(x_\mu)^2(1-F(x_\mu))^2\sum_{k=0}^{K}\left(-\frac{1}{p_{k0}} - \frac{1}{p_{k1}}\right) \\ &\quad \left[H(x_{\alpha_k}, x_\mu) - H(x_{\alpha_{k+1}}, x_\mu)\right]\left[F(x_{\alpha_k})(1-F(x_{\alpha_k})) - F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\right] \end{aligned}$$

$$\begin{aligned} \boldsymbol{M}_{\alpha_j\tau} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial \alpha_j}\frac{\partial p_{k0}}{\partial \tau} + \frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial \alpha_j}\frac{\partial p_{k1}}{\partial \tau}\right] \\ &= \frac{1}{\beta}F(x_\mu)(1-F(x_\mu))F(x_{\alpha_j})(1-F(x_{\alpha_j})) \\ &\quad \left[(1-H(x_\mu, x_{\alpha_j}))\left(-\frac{1}{p_{(j-1)0}}\left[F(x_{\alpha_{j-1}})(1-F(x_{\alpha_{j-1}})) - F(x_{\alpha_j})(1-F(x_{\alpha_j}))\right]\right.\right. \\ &\quad \left.+\frac{1}{p_{j0}}\left[F(x_{\alpha_j})(1-F(x_{\alpha_j})) - F(x_{\alpha_{j+1}})(1-F(x_{\alpha_{j+1}}))\right]\right) \\ &\quad +H(x_\mu, x_{\alpha_j})\left(\frac{1}{p_{(j-1)1}}\left[F(x_{\alpha_{j-1}})(1-F(x_{\alpha_{j-1}})) - F(x_{\alpha_j})(1-F(x_{\alpha_j}))\right]\right. \\ &\quad \left.\left.-\frac{1}{p_{j1}}\left[F(x_{\alpha_j})(1-F(x_{\alpha_j})) - F(x_{\alpha_{j+1}})(1-F(x_{\alpha_{j+1}}))\right]\right)\right] \end{aligned}$$

$$
\begin{aligned}
\boldsymbol{M}_{\sigma\tau} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial\sigma}\frac{\partial p_{k0}}{\partial\tau}+\frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial\sigma}\frac{\partial p_{k1}}{\partial\tau}\right] \\
&= \frac{1}{\sigma}x_\mu F(x_\mu)^2(1-F(x_\mu))^2\sum_{k=0}^{K}\left(-\frac{1}{p_{k0}}-\frac{1}{p_{k1}}\right) \\
&\quad \left[H(x_{\alpha_k},x_\mu)-H(x_{\alpha_{k+1}},x_\mu)\right]\left[F(x_{\alpha_k})(1-F(x_{\alpha_k}))-F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\right] \\
&= x_\mu \boldsymbol{M}_{\mu\tau}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{M}_{\beta\tau} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\frac{\partial p_{k0}}{\partial\beta}\frac{\partial p_{k0}}{\partial\tau}+\frac{1}{p_{k1}}\frac{\partial p_{k1}}{\partial\beta}\frac{\partial p_{k1}}{\partial\tau}\right] \\
&= F(x_\mu)(1-F(x_\mu))\sum_{k=0}^{K}\left[F(x_{\alpha_k})(1-F(x_{\alpha_k}))-F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\right] \\
&\quad \left(-\frac{1}{p_{k0}}\left[-\frac{1}{\beta}x_{\alpha_k}F(x_{\alpha_k})(1-F(x_{\alpha_k}))(1-H(x_\mu,x_{\alpha_k}))\right.\right. \\
&\quad \left.+\frac{1}{\beta}x_{\alpha_{k+1}}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))(1-H(x_\mu,x_{\alpha_{k+1}}))\right] \\
&\quad +\frac{1}{p_{k1}}\left[-\frac{1}{\beta}x_{\alpha_k}F(x_{\alpha_k})(1-F(x_{\alpha_k}))H(x_\mu,x_{\alpha_k})\right. \\
&\quad \left.\left.+\frac{1}{\beta}x_{\alpha_{k+1}}F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))H(x_\mu,x_{\alpha_{k+1}})\right]\right) \\
&= \sum_{k=0}^{K}x_{\alpha_k}\boldsymbol{M}_{\alpha_k\tau}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{M}_{\tau\tau} &= \sum_{k=0}^{K}\left[\frac{1}{p_{k0}}\left(\frac{\partial p_{k0}}{\partial\tau}\right)^2+\frac{1}{p_{k1}}\left(\frac{\partial p_{k1}}{\partial\tau}\right)^2\right] \\
&= F(x_\mu)^2(1-F(x_\mu))^2\sum_{k=0}^{K}\left(\frac{1}{p_{k0}}+\frac{1}{p_{k1}}\right) \\
&\quad \left[F(x_{\alpha_k})(1-F(x_{\alpha_k}))-F(x_{\alpha_{k+1}})(1-F(x_{\alpha_{k+1}}))\right]^2
\end{aligned}
$$

Now it can be seen quite easily that the information matrix is the one given in Chapter 5.

B Figures

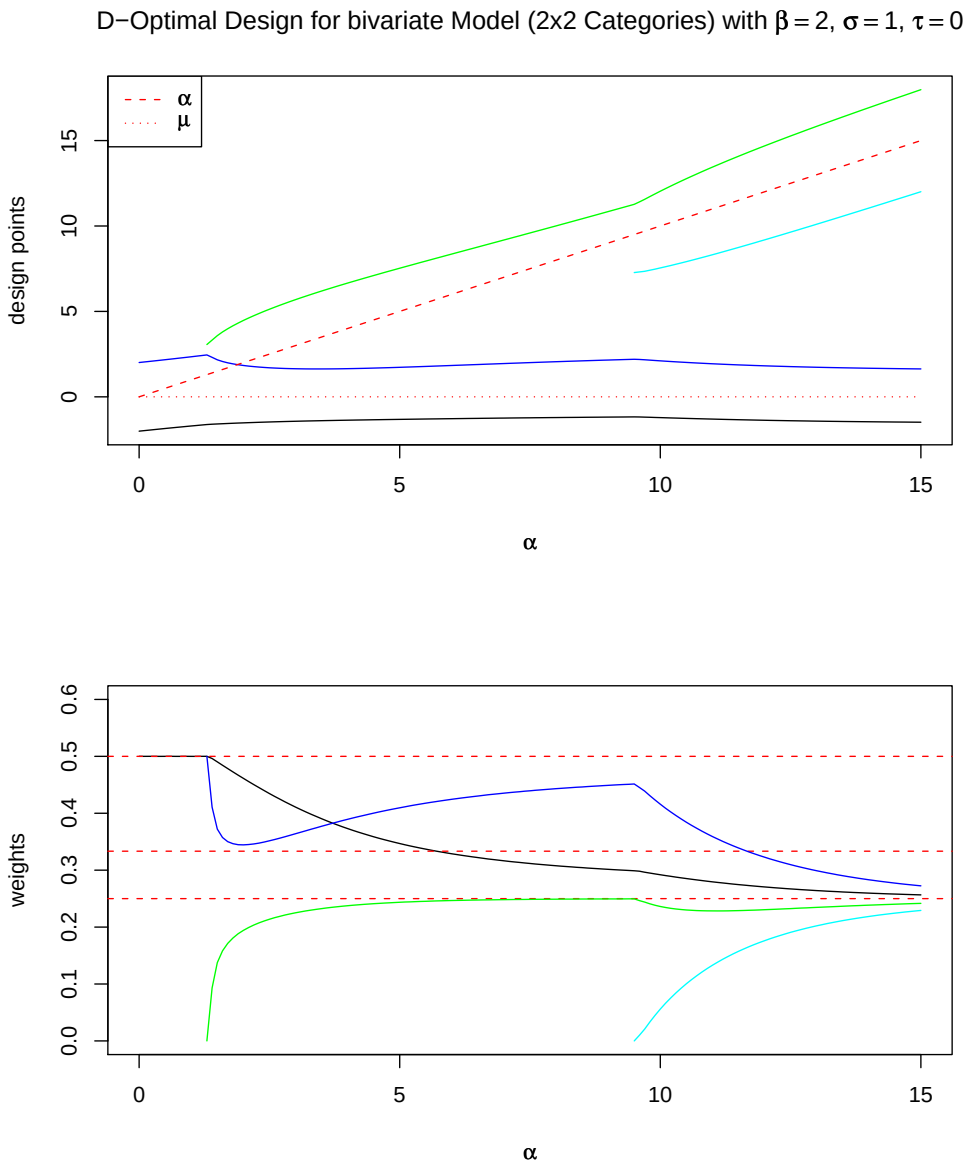


Figure B.1: D-optimal design for the bivariate model with $\beta = 2, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

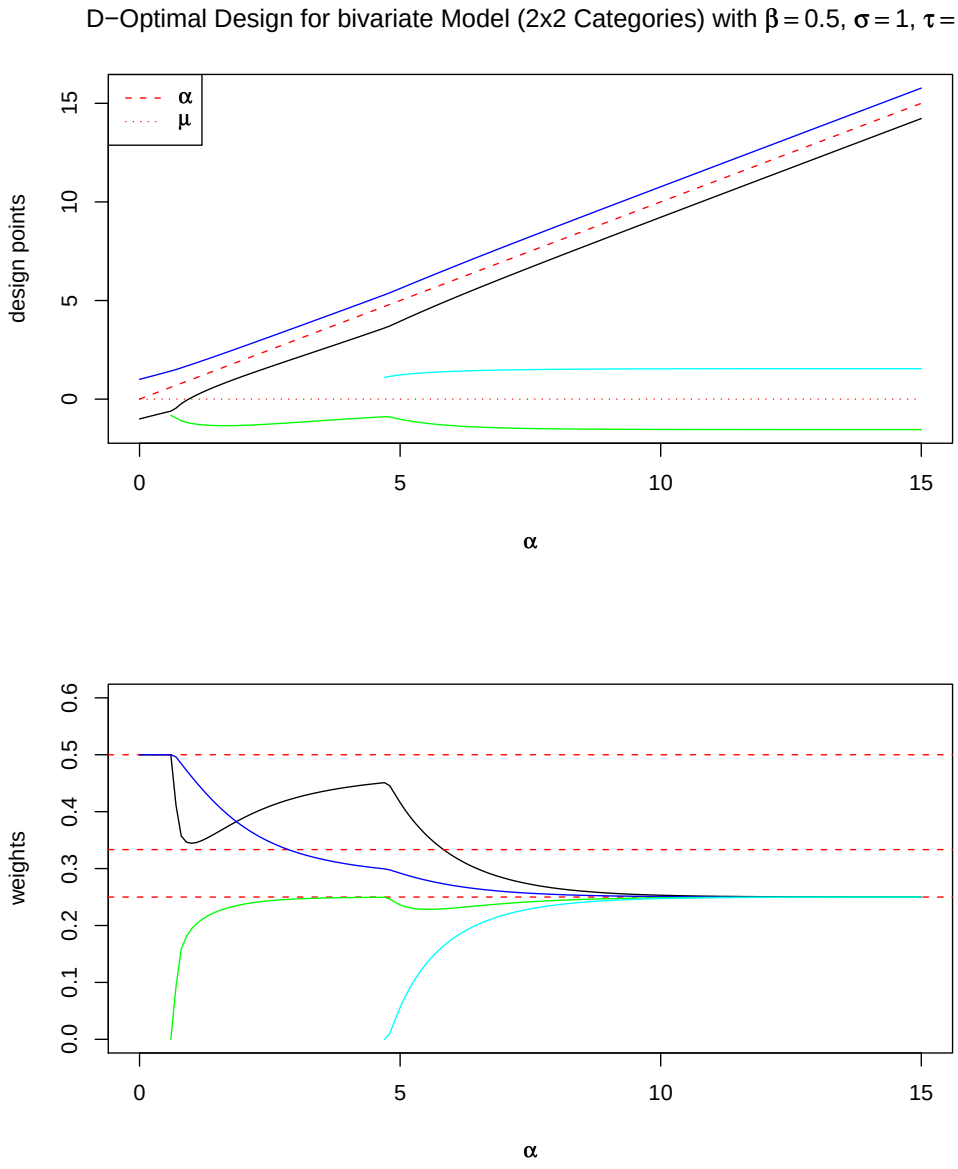


Figure B.2: D-optimal design for the bivariate model with $\beta = 0.5, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

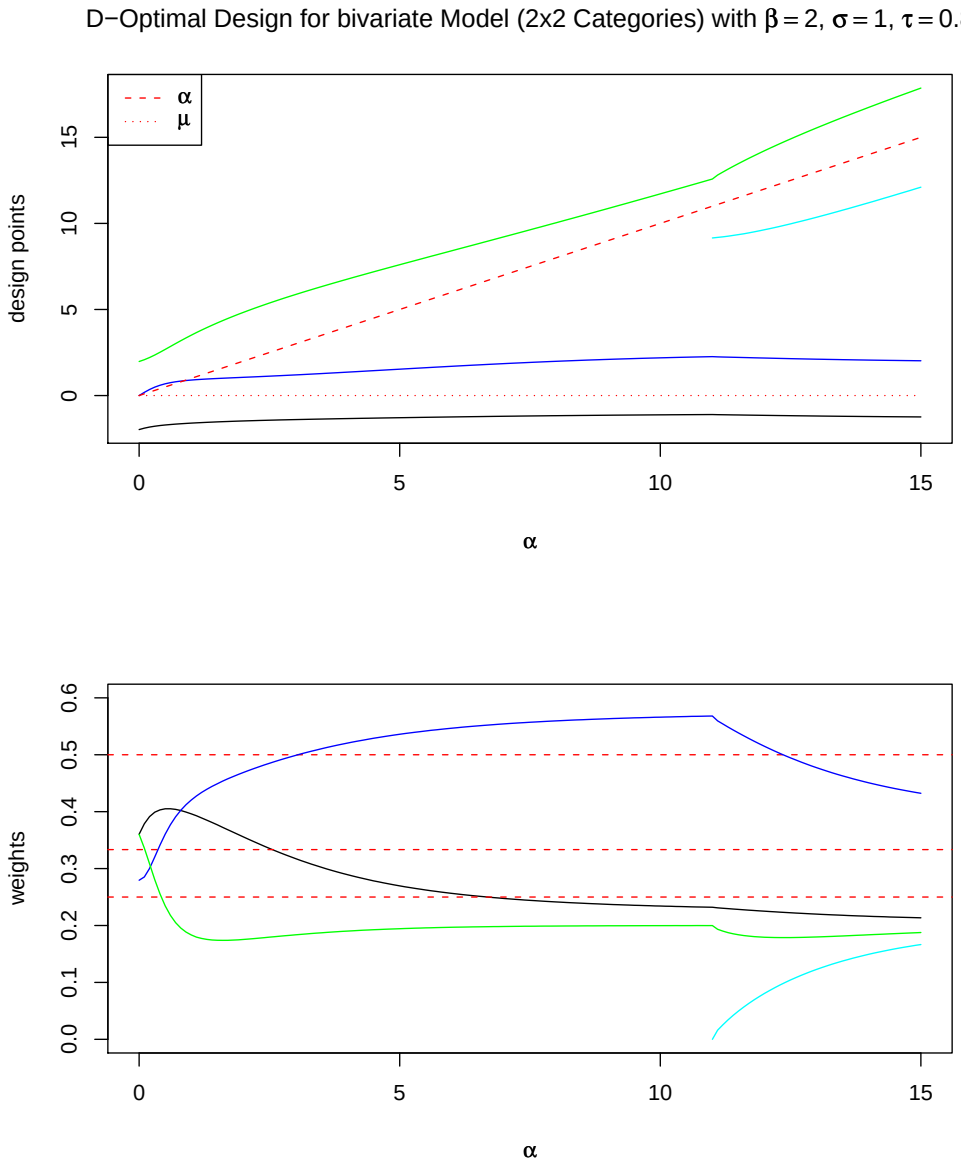


Figure B.3: D-optimal design for the bivariate model with $\beta = 2, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

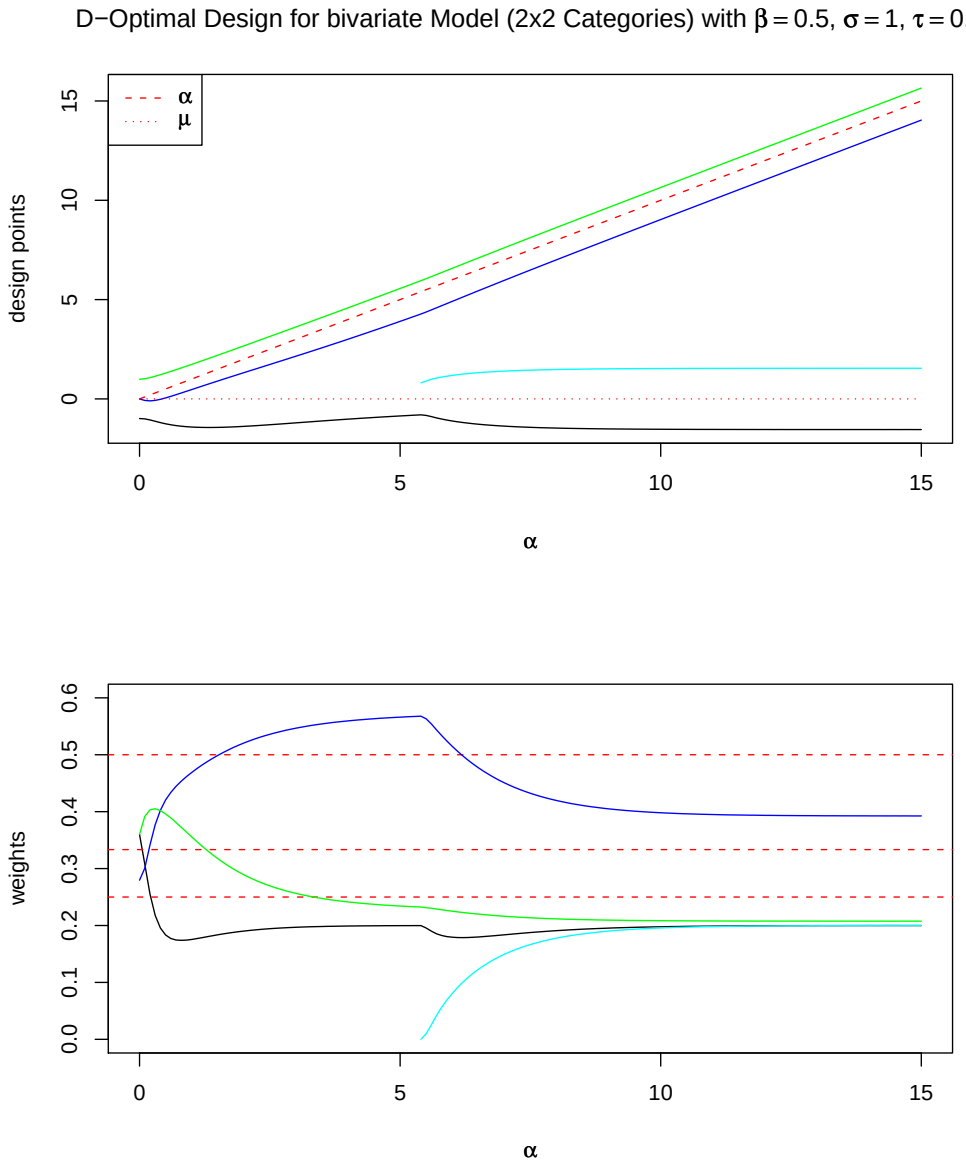


Figure B.4: D-optimal design for the bivariate model with $\beta = 0.5, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

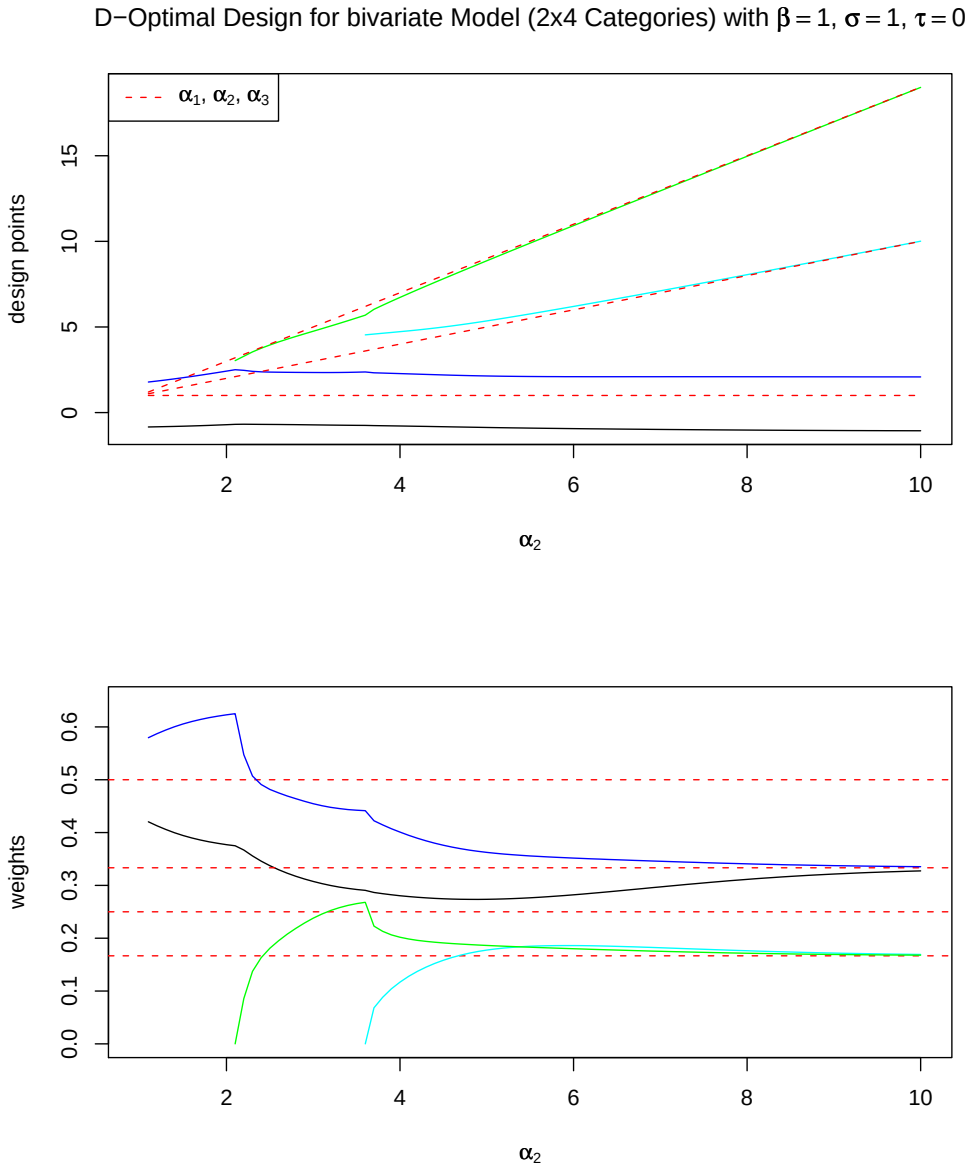


Figure B.5: D-optimal design for the bivariate model with $\alpha_1 = 1, \beta = 1, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

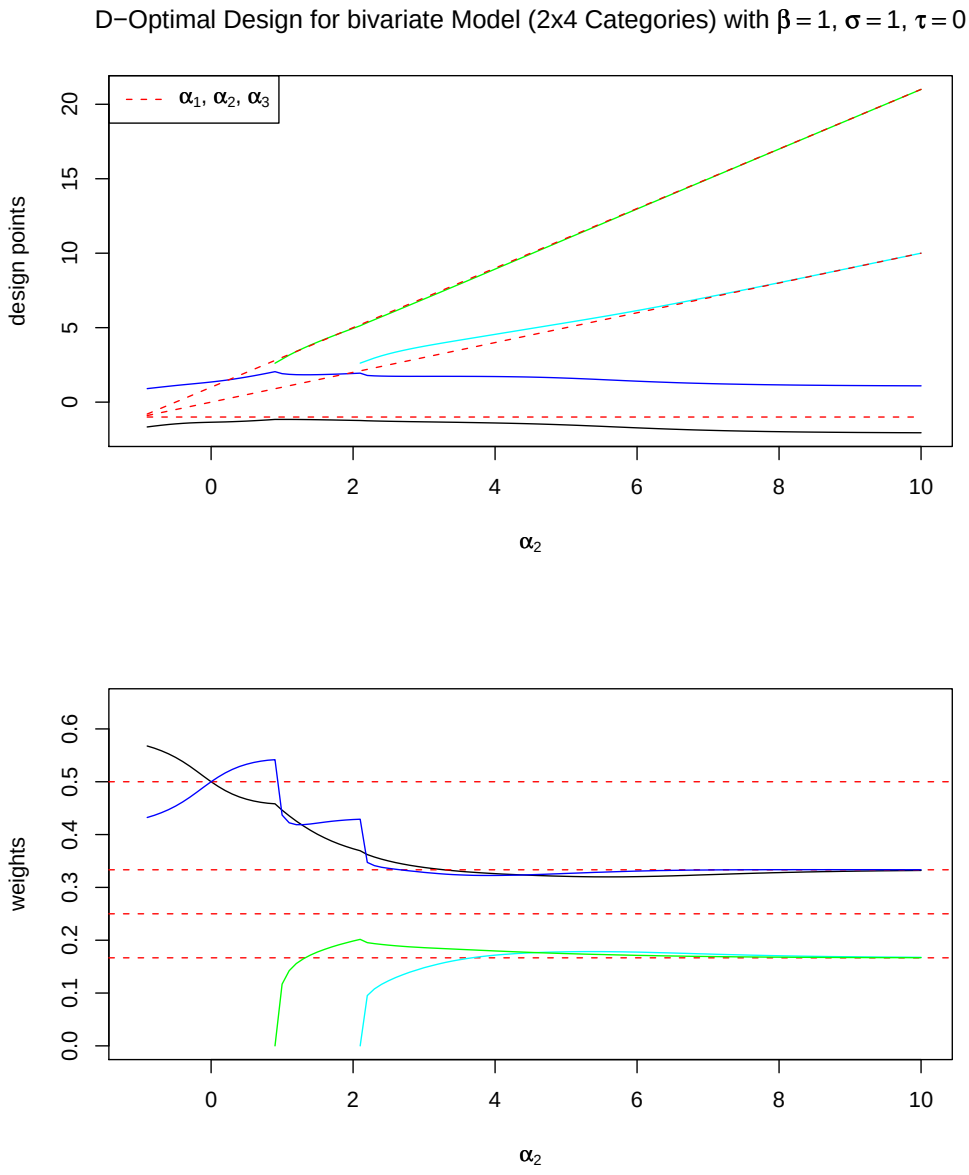


Figure B.6: D-optimal design for the bivariate model with $\alpha_1 = -1, \beta = 1, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

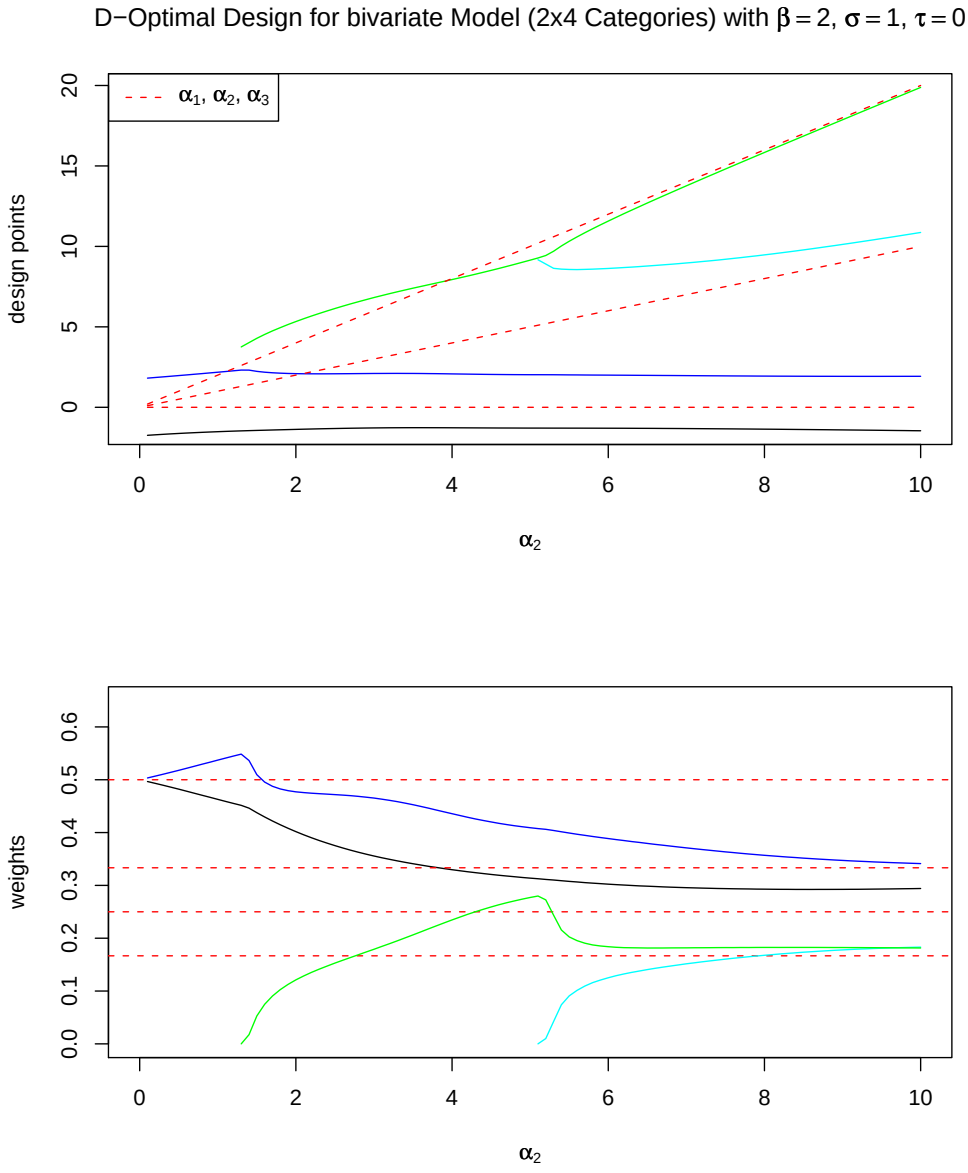


Figure B.7: D-optimal design for the bivariate model with $\alpha_1 = 0, \beta = 2, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

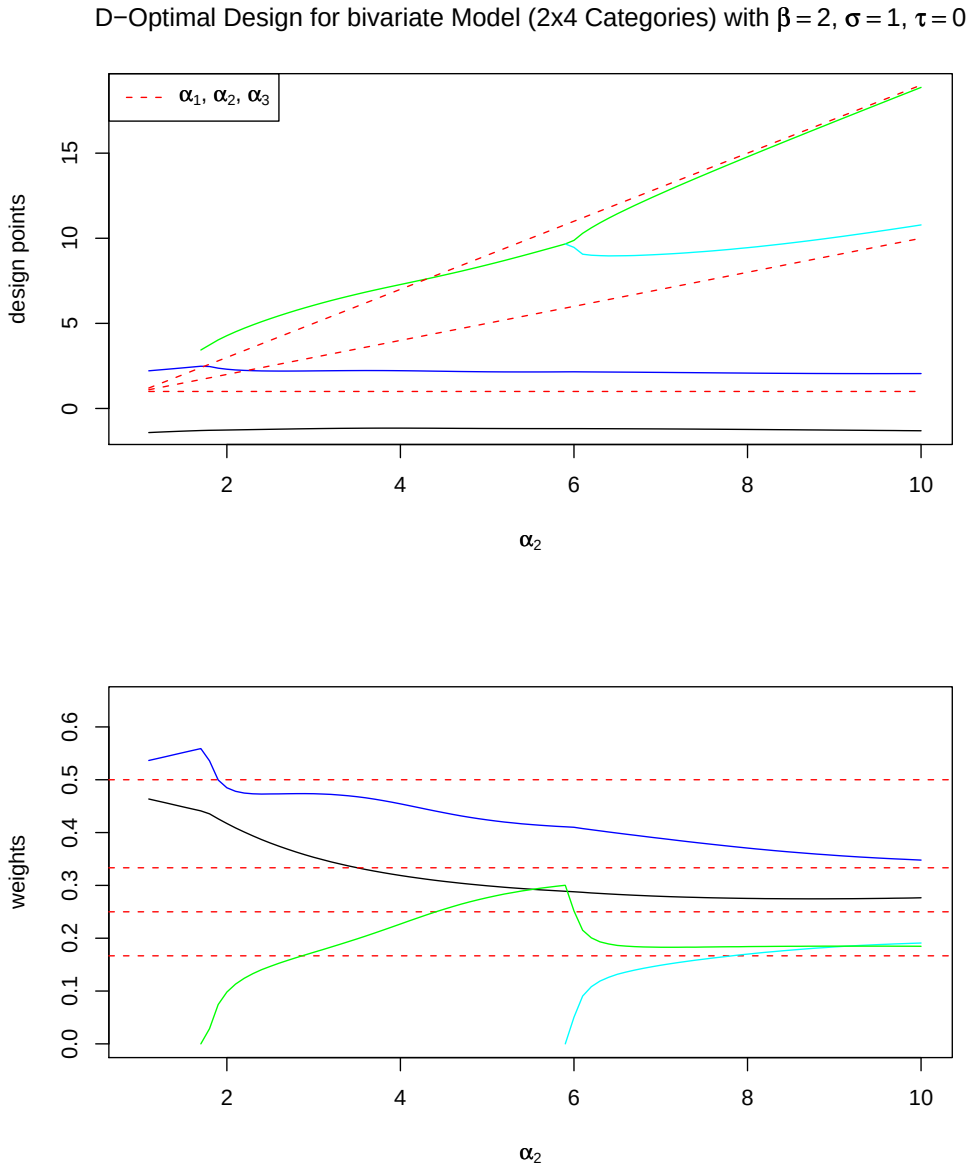


Figure B.8: D-optimal design for the bivariate model with $\alpha_1 = 1, \beta = 2, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

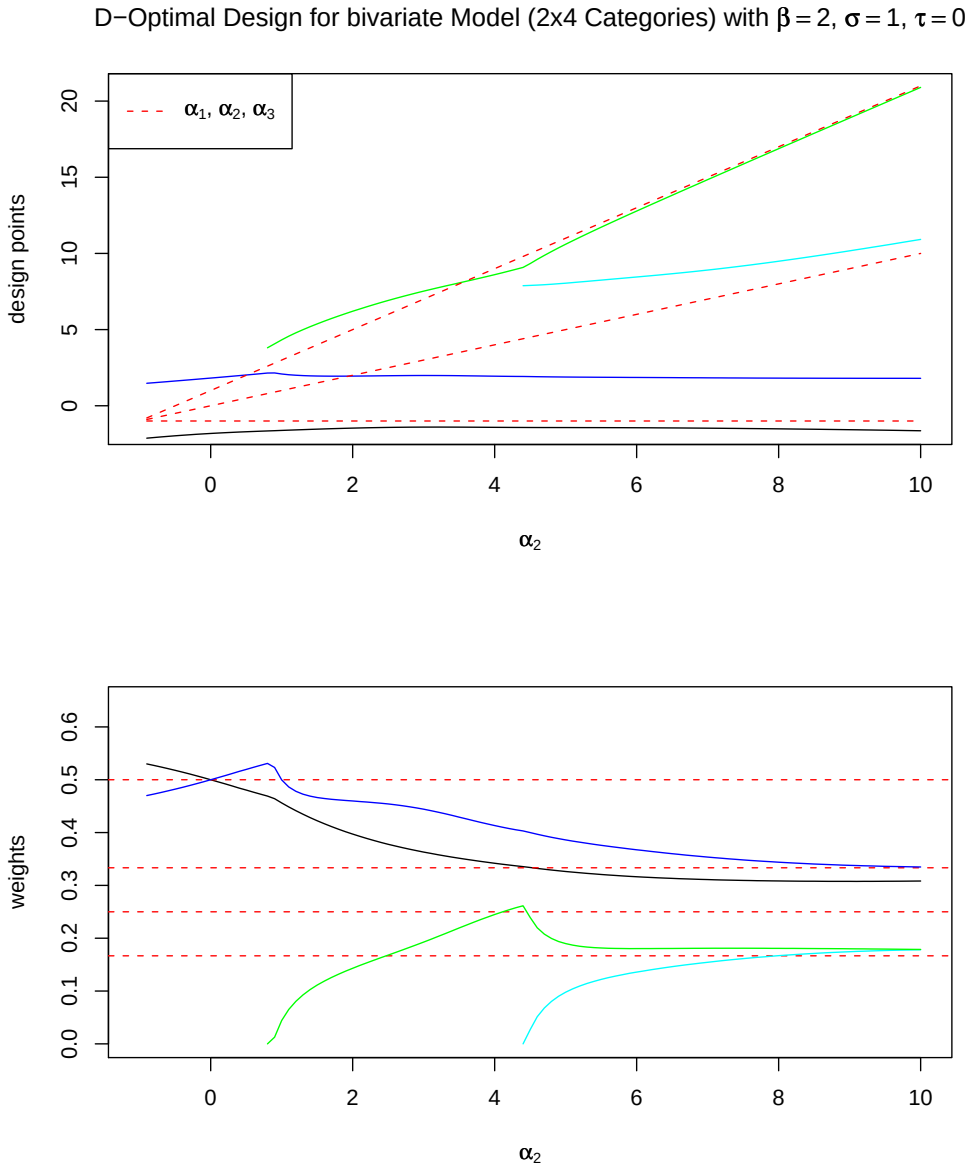


Figure B.9: D-optimal design for the bivariate model with $\alpha_1 = -1, \beta = 2, \sigma = 1$ and $\tau = 0$; top: optimal design points, bottom: optimal weights.

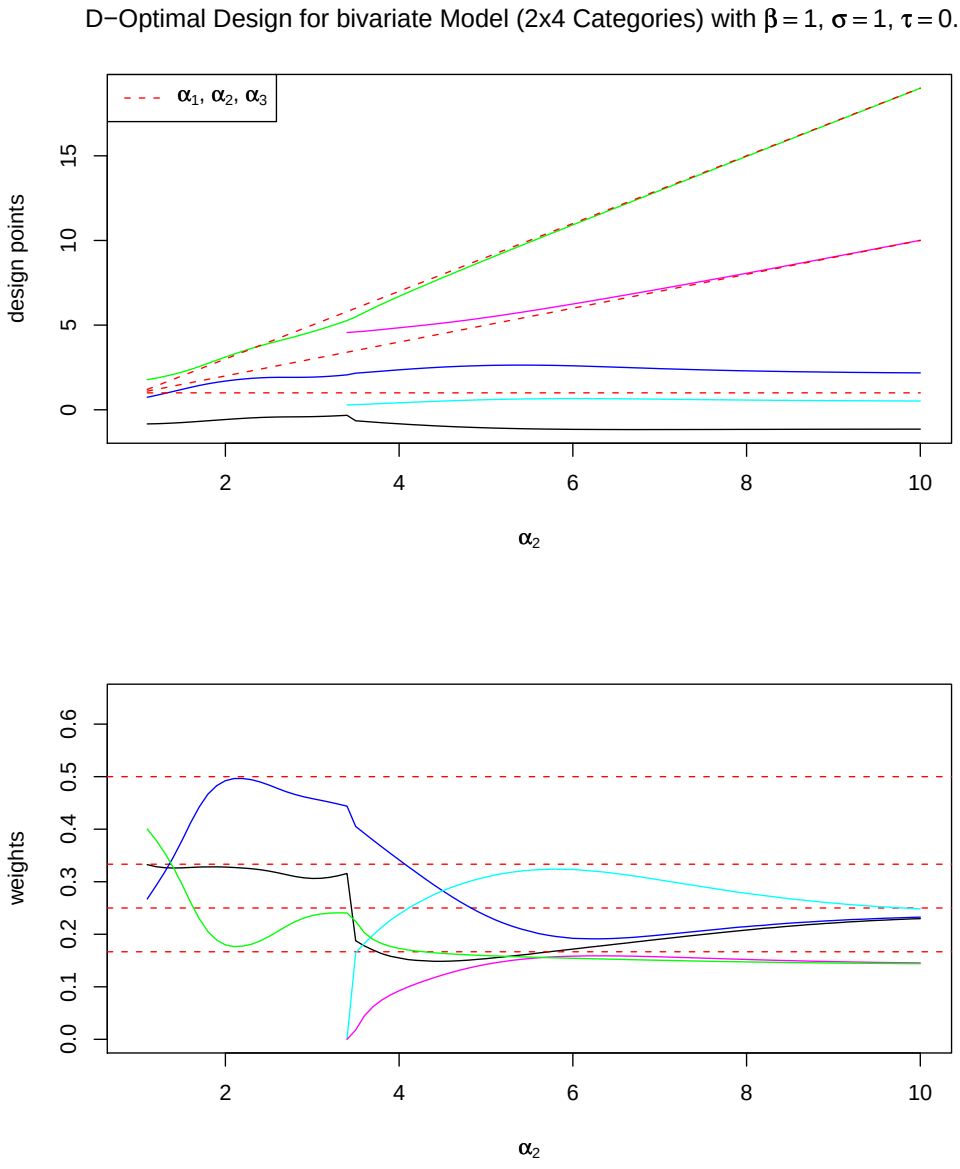


Figure B.10: D-optimal design for the bivariate model with $\alpha_1 = 1, \beta = 1, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

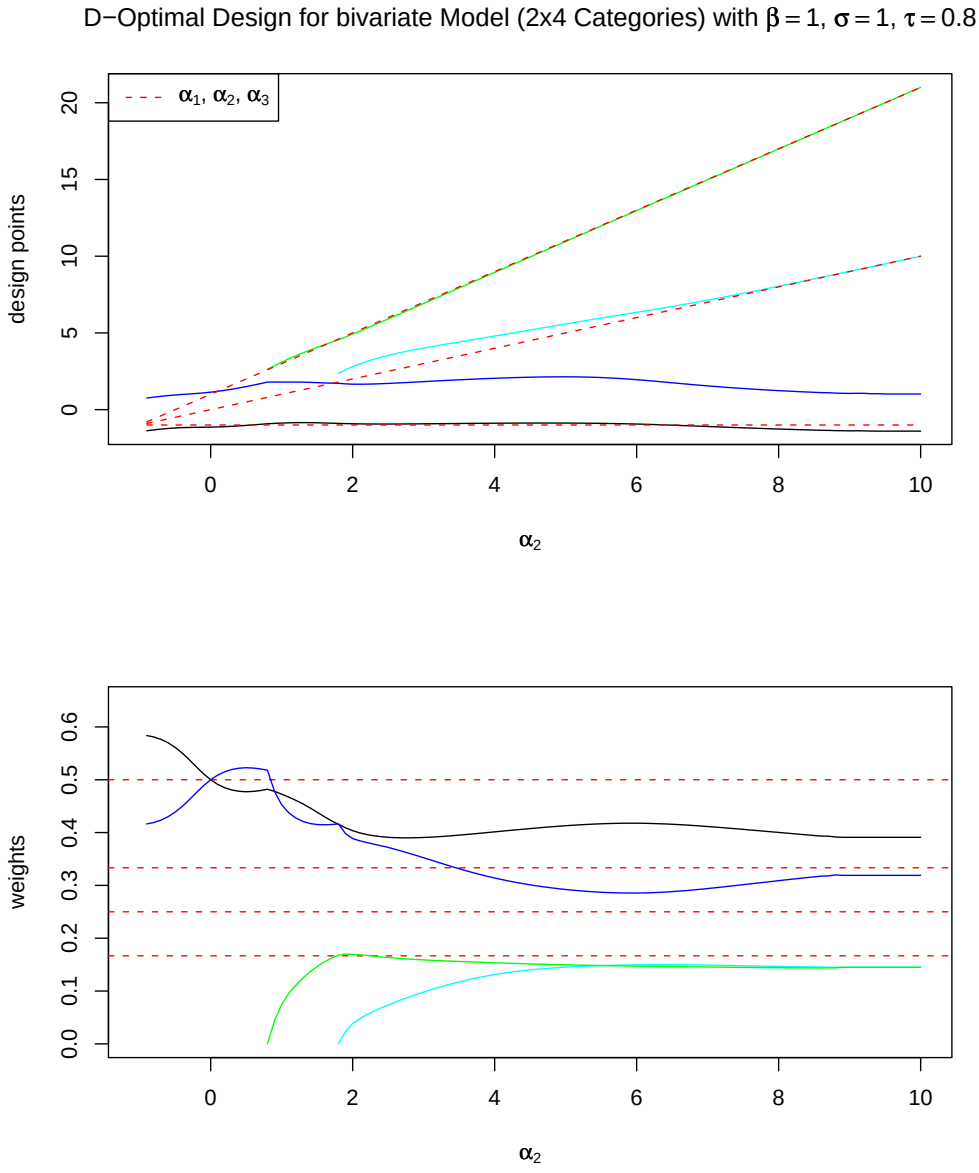


Figure B.11: D-optimal design for the bivariate model with $\alpha_1 = -1, \beta = 1, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

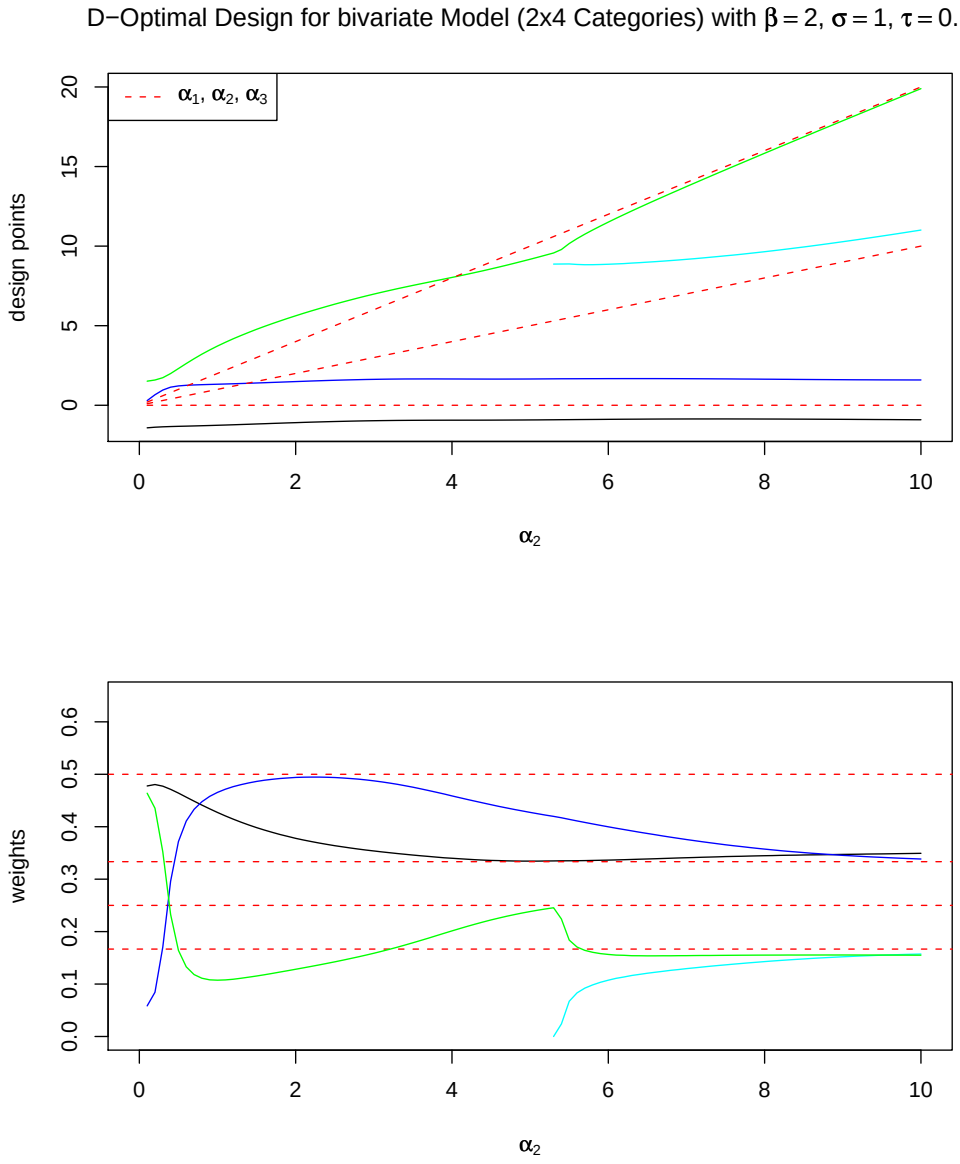


Figure B.12: D-optimal design for the bivariate model with $\alpha_1 = 0, \beta = 2, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

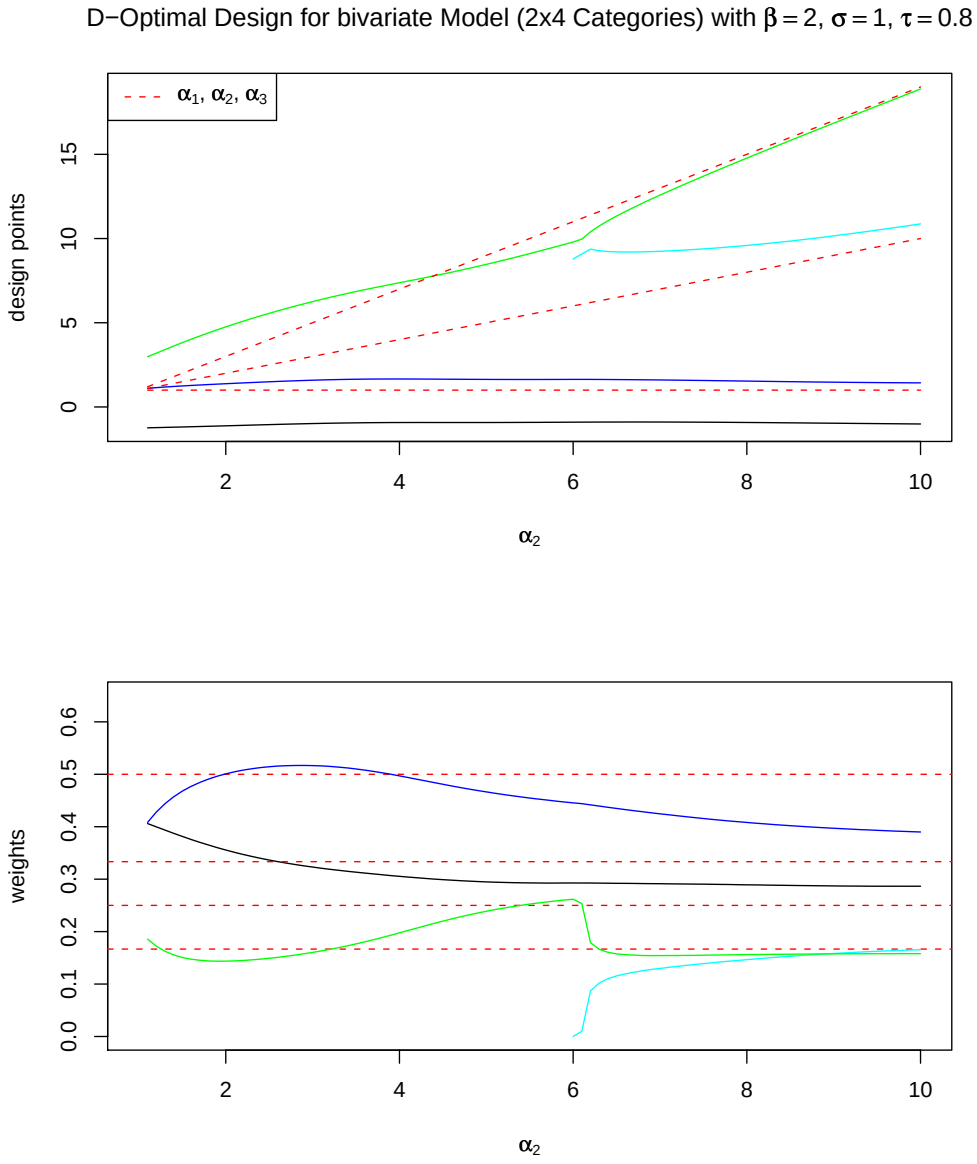


Figure B.13: D-optimal design for the bivariate model with $\alpha_1 = 1, \beta = 2, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

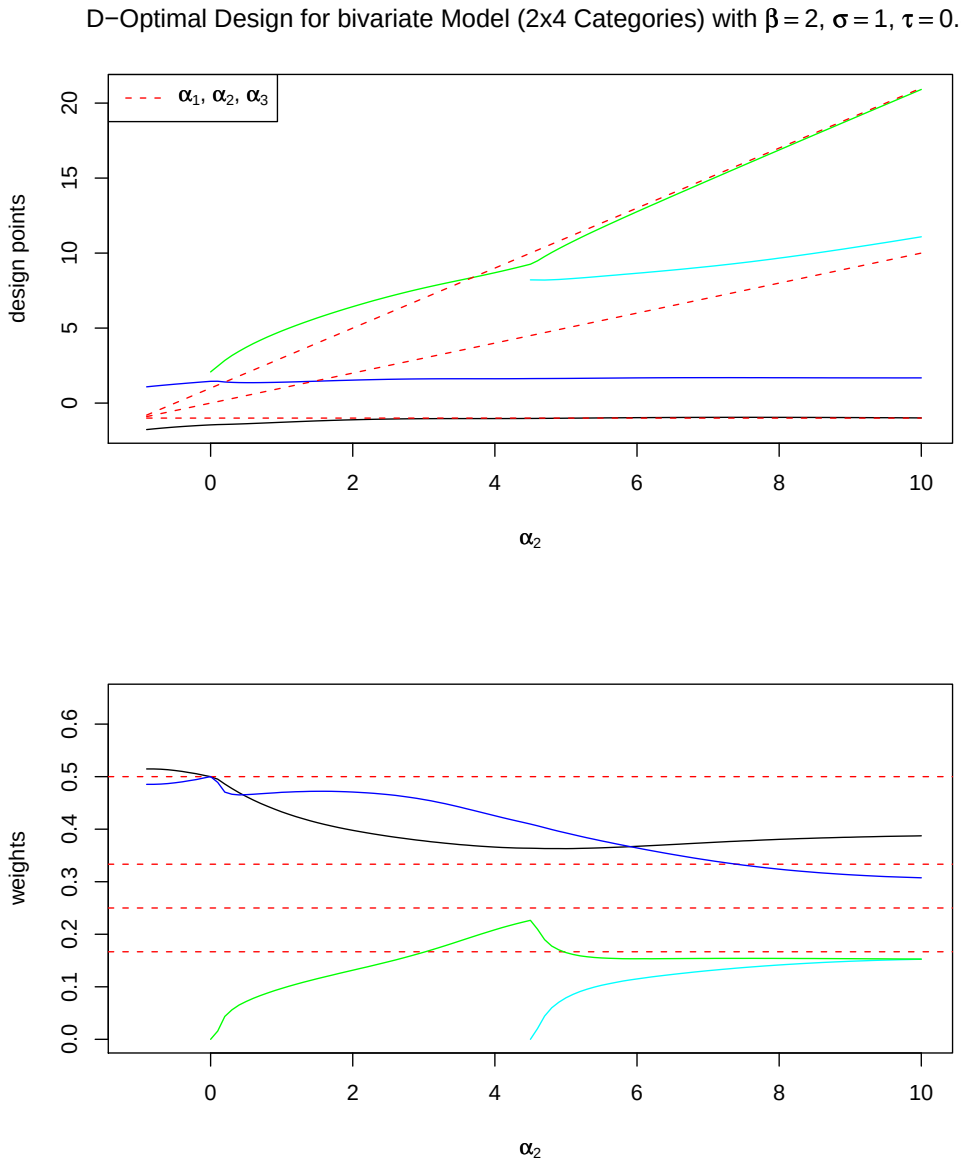


Figure B.14: D-optimal design for the bivariate model with $\alpha_1 = -1, \beta = 2, \sigma = 1$ and $\tau = 0.8$; top: optimal design points, bottom: optimal weights.

C Tables

C.1 Scenario II

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
0.6	3.78	0.00	0.00	0.00	0.00
1.2	4.29	0.00	0.00	0.00	0.00
2.0	5.04	0.09	0.28	0.09	0.24
3.0	5.82	0.03	0.04	0.02	0.02
4.0	6.99	0.14	0.15	0.06	0.07
5.3	8.76	0.61	0.71	0.28	0.29
7.0	11.73	2.99	3.42	1.28	1.46
9.3	15.73	13.77	15.29	6.39	6.92
12.4	18.17	34.40	33.87	25.12	22.83
16.5	**12.77**	**40.55**	**39.05**	**47.84**	**49.51**
22.0	3.25	7.40	7.11	18.40	17.72
29.4	0.20	0.02	0.08	0.53	0.94
none	3.47	0.00	0.00	0.00	0.00

Table C.1: Percentage of each dose being estimated as the MTD in **Scenario II** for the **3+3 design** and different settings in **Bayesian ADEPT**; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	2.82	3.41	2.68	3.41	2.75	3.43	2.67	2.67
1.2	3.15	3.17	3.21	3.17	3.10	3.04	3.32	3.42
2.0	3.19	3.35	3.45	3.35	3.32	3.52	3.54	3.63
3.0	5.28	5.30	5.36	5.30	5.23	5.27	5.28	5.99
4.0	5.64	5.46	6.63	5.46	5.69	5.50	6.48	6.81
5.3	6.01	6.04	9.29	6.04	6.00	6.15	9.32	11.68
7.0	6.19	6.51	17.28	6.51	6.44	6.91	17.55	19.06
9.3	9.36	10.71	20.74	10.71	9.89	11.38	20.95	21.57
12.4	20.33	19.90	16.13	19.90	20.12	19.64	16.24	13.48
16.5	**22.91**	**19.56**	**8.66**	**19.56**	**22.91**	**18.97**	**8.73**	**6.94**
22.0	7.94	8.58	2.14	8.58	7.38	8.12	2.14	1.01
29.4	3.65	4.57	0.32	4.57	3.66	4.57	0.32	0.23
none	3.53	3.45	3.42	3.45	3.52	3.50	3.46	3.53

Table C.2: Percentage of each dose being estimated as the MTD in **Scenario II** for **SLOD based on the logistic model** with different optimality criteria, design regions, and cohort sizes n; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	3.73	3.45	3.66	3.15	3.67	3.61	3.65	3.59
1.2	3.25	3.32	3.23	3.57	3.25	3.57	3.23	3.46
2.0	3.25	3.14	3.22	3.38	3.30	3.06	3.26	3.07
3.0	3.35	3.25	3.41	3.60	3.32	3.65	3.36	3.58
4.0	3.78	3.65	3.92	3.62	3.81	3.79	3.88	3.69
5.3	3.48	3.57	3.83	3.57	3.46	3.38	3.84	3.59
7.0	4.52	4.44	5.23	5.12	4.44	4.39	5.19	5.06
9.3	8.85	9.23	10.23	10.01	8.65	9.02	10.14	9.61
12.4	20.20	19.29	21.28	20.39	19.94	18.64	21.14	20.67
16.5	**26.70**	**23.02**	**24.93**	**25.99**	**27.01**	**23.77**	**25.18**	**25.66**
22.0	10.15	11.29	9.62	10.01	10.43	10.93	9.66	9.93
29.4	5.40	8.75	4.11	4.14	5.39	8.67	4.14	4.51
none	3.34	3.60	3.33	3.45	3.33	3.52	3.33	3.58

Table C.3: Percentage of each dose being estimated as the MTD in **Scenario II** for **SLOD based on the 4 category proportional odds model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	2.84	3.18	2.76	2.97	2.83	3.25	2.73	2.76
1.2	2.84	2.88	2.79	2.82	2.73	2.78	2.91	2.95
2.0	2.73	2.94	2.76	3.20	2.76	3.12	2.82	2.86
3.0	2.54	2.68	2.75	2.85	2.46	2.20	2.63	3.17
4.0	2.03	2.13	3.19	3.31	2.08	2.43	3.19	3.67
5.3	2.39	3.06	6.50	8.58	2.32	2.96	6.42	8.19
7.0	2.65	2.69	15.56	17.69	3.04	3.73	16.10	18.57
9.3	7.50	7.79	22.90	25.09	7.53	8.83	23.39	25.53
12.4	25.17	27.44	21.52	19.09	25.62	26.77	21.07	17.99
16.5	**30.35**	**24.99**	**11.39**	**9.32**	**30.29**	**24.37**	**11.06**	**9.10**
22.0	11.71	11.84	3.96	1.42	11.16	11.30	3.83	1.36
29.4	3.86	4.78	0.55	0.29	3.81	4.95	0.48	0.26
none	3.39	3.60	3.37	3.37	3.37	3.31	3.37	3.59

Table C.4: Percentage of each dose being estimated as the MTD in **Scenario II** for **SLOD based on the bivariate model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	**3+3 Design**	**Bayesian ADEPT**			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
$\overline{N}$	31.30	24.91	26.94	29.09	30.43
$\overline{N_{DLT}}$	3.77	7.37	7.91	8.46	8.70
$\overline{N_{>MTD}}$	1.01	7.78	9.12	8.56	8.50
MSE	89.37	19.27	20.93	15.84	16.72

Table C.5: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), and mean squared error (MSE) in **Scenario II** for the **3+3 design** and different settings in **Bayesian ADEPT**.

	D-criterion				**c-criterion**			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	28.30	28.78	28.30	28.78	28.25	28.71	28.30	28.69
$\overline{N_{DLT}}$	4.35	3.48	4.08	3.48	4.10	3.39	4.09	4.12
$\overline{N_{>MTD}}$	3.10	1.22	2.52	1.22	2.66	1.00	2.53	2.44
MSE	72.37	76.76	86.97	76.76	72.60	77.61	87.49	93.71
CI	2.42	2.72	2.16	2.72	2.35	2.35	2.16	2.16

Table C.6: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of the 95% confidence interval for the MTD (CI) in **Scenario II** for different settings of **SLOD based on the logistic model**.

	D-criterion				**c-criterion**			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	30.16	30.54	30.16	30.53	30.16	30.42	30.16	30.43
$\overline{N_{DLT}}$	5.82	5.19	5.56	5.66	5.83	5.17	5.64	5.71
$\overline{N_{>MTD}}$	3.67	2.46	2.77	2.91	3.58	2.41	2.82	2.98
MSE	66.81	71.98	66.53	65.78	66.56	72.99	66.40	66.57
CI	2.57	2.88	2.56	2.57	2.55	2.55	2.56	2.56

Table C.7: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of the 95% confidence interval for the MTD (CI) in **Scenario II** for different settings of **SLOD based on the 4 category proportional odds model**.

	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	30.67	31.05	30.67	31.04	30.67	31.12	30.67	31.04
$\overline{N_{DLT}}$	5.41	4.28	4.68	4.67	5.14	4.15	4.70	4.67
$\overline{N_{>MTD}}$	4.67	2.25	2.97	2.85	4.20	1.97	3.00	2.85
MSE MTD	52.80	57.89	71.92	78.13	52.61	59.05	72.47	78.81
CI MTD	2.62	2.97	2.40	2.39	2.55	2.55	2.39	2.39
MSE minED	16.65	17.33	17.52	17.74	16.72	17.56	17.55	17.69
CI minED	2.71	2.84	2.84	2.84	2.67	2.67	2.85	2.85

Table C.8: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error for the MTD (MSE MTD) and minimum effective dose (MSE minED), and median width of the 95% confidence intervals for the MTD (CI MTD) and the minimum effective dose (CI minED) in **Scenario II** for different settings of **SLOD based on the bivariate model**.

C.2 Scenario III

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
0.6	1.18	0.00	0.00	0.00	0.00
1.2	1.29	0.00	0.00	0.00	0.00
2.0	1.41	4.09	2.86	11.22	6.60
3.0	1.66	0.00	0.00	0.00	0.00
4.0	1.87	0.00	0.00	0.00	0.00
5.3	2.26	0.00	0.00	0.00	0.00
7.0	2.93	0.01	0.02	0.01	0.02
9.3	4.27	0.08	0.09	0.05	0.05
12.4	6.37	0.35	0.29	0.18	0.18
16.5	10.45	1.88	1.51	0.65	1.19
22.0	18.00	2.29	2.09	1.89	1.88
29.4	**47.19**	**91.30**	**93.15**	**86.00**	**90.07**
none	1.12	0.00	0.00	0.00	0.00

Table C.9: Percentage of each dose being estimated as the MTD in **Scenario III** for the **3+3 design** and different settings in **Bayesian ADEPT**; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	1.09	1.33	1.00	0.98	1.09	1.33	1.00	1.01
1.2	1.05	1.09	1.14	1.10	1.02	1.02	1.17	1.15
2.0	1.07	1.29	1.09	1.28	1.10	1.32	1.10	1.27
3.0	1.94	2.00	1.93	2.24	1.90	1.95	1.93	2.23
4.0	2.30	2.10	2.15	2.25	2.30	2.08	2.13	2.21
5.3	2.40	3.71	2.87	3.60	2.44	3.75	2.85	3.61
7.0	2.79	2.85	3.40	8.06	2.81	2.91	3.40	8.06
9.3	3.50	3.01	5.81	9.95	3.50	3.10	5.82	9.92
12.4	4.32	5.38	17.15	13.39	4.37	5.51	17.15	13.47
16.5	6.87	6.97	18.99	15.35	7.10	7.37	18.98	15.30
22.0	14.49	16.02	16.18	13.24	14.44	15.51	16.19	13.16
29.4	**57.04**	**53.07**	**27.22**	**27.36**	**56.80**	**52.98**	**27.20**	**27.45**
none	1.11	1.18	1.09	1.19	1.13	1.18	1.09	1.16

Table C.10: Percentage of each dose being estimated as the MTD in **Scenario III** for **SLOD based on the logistic model** with different optimality criteria, design regions, and cohort sizes n; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	1.32	1.20	1.29	1.11	1.32	1.19	1.29	1.22
1.2	1.11	1.54	1.13	1.51	1.10	1.57	1.13	1.54
2.0	1.13	1.21	1.15	1.23	1.13	1.16	1.14	1.11
3.0	1.28	1.32	1.25	1.19	1.28	1.28	1.25	1.32
4.0	0.96	1.08	0.95	1.18	0.97	1.14	0.96	1.06
5.3	1.20	1.07	1.19	1.03	1.20	1.11	1.20	1.09
7.0	1.10	1.06	1.09	1.01	1.10	1.08	1.09	1.10
9.3	1.07	1.48	1.09	1.47	1.03	1.48	1.09	1.38
12.4	1.56	1.73	1.78	1.72	1.56	1.84	1.77	1.92
16.5	2.90	3.17	3.71	3.37	2.76	3.15	3.68	3.14
22.0	11.16	11.98	10.67	11.10	11.08	11.77	10.63	10.81
29.4	**74.25**	**72.20**	**73.74**	**73.06**	**74.51**	**72.26**	**73.81**	**73.43**
none	0.96	0.96	0.96	1.02	0.96	0.97	0.96	0.88

Table C.11: Percentage of each dose being estimated as the MTD in **Scenario III** for **SLOD based on the 4 category proportional odds model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	1.00	1.34	0.89	0.86	1.00	1.30	0.89	0.95
1.2	1.02	1.09	1.07	0.93	0.97	0.83	1.08	1.20
2.0	1.00	1.20	1.00	1.03	1.00	1.22	1.01	1.28
3.0	0.98	0.91	0.97	1.22	0.96	0.96	0.98	1.18
4.0	1.13	0.92	1.15	1.05	1.11	1.11	1.18	1.05
5.3	1.17	2.29	1.71	2.27	1.18	2.42	1.82	2.45
7.0	1.40	1.64	1.73	4.71	1.36	1.41	1.96	6.52
9.3	1.53	1.58	3.43	7.24	1.59	1.63	4.23	8.22
12.4	1.90	3.48	12.49	12.38	1.93	3.33	15.82	12.22
16.5	4.01	4.23	19.88	18.04	4.53	4.85	18.82	16.82
22.0	15.89	17.80	19.63	17.72	15.34	17.55	17.95	16.00
29.4	**67.98**	**62.50**	**35.06**	**31.58**	**68.04**	**62.38**	**33.27**	**31.06**
none	0.99	1.02	0.99	0.97	0.99	1.01	0.99	1.05

Table C.12: Percentage of each dose being estimated as the MTD in **Scenario III** for **SLOD based on the bivariate model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
$\overline{N}$	38.11	58.20	58.47	58.96	58.56
$\overline{N_{DLT}}$	2.46	7.37	7.48	10.62	10.26
MSE	146.54	36.48	26.41	87.13	53.41

Table C.13: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), and mean squared error (MSE) in **Scenario III** for the **3+3 design** and different settings in **Bayesian ADEPT**.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	35.59	36.05	35.69	36.02	35.57	36.04	35.69	36.01
$\overline{N_{DLT}}$	2.53	2.33	2.69	2.68	2.52	2.31	2.69	2.68
MSE	129.22	142.39	202.03	232.41	129.68	143.22	202.17	232.55
CI	3.88	4.54	4.13	4.01	3.88	3.88	4.13	4.13

Table C.14: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), mean squared error (MSE), and median width of the 95% confidence interval for the MTD (CI) in **Scenario III** for different settings of **SLOD based on the logistic model**.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	37.21	37.60	37.21	37.60	37.21	37.59	37.21	37.57
$\overline{N_{DLT}}$	3.51	3.34	3.64	3.69	3.51	3.34	3.65	3.66
MSE	76.34	82.57	77.82	80.80	75.89	82.97	77.77	81.55
CI	4.09	4.44	3.91	3.94	4.10	4.10	3.92	3.92

Table C.15: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), mean squared error (MSE), and median width of the 95% confidence interval for the MTD (CI) in **Scenario III** for different settings of **SLOD based on the 4 category proportional odds model**.

Dose	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	37.62	38.01	37.62	38.04	37.62	37.98	37.62	37.92
$\overline{N_{DLT}}$	2.84	2.60	2.94	2.93	2.86	2.57	2.99	2.93
MSE MTD	79.66	96.78	151.32	180.72	79.74	96.35	163.84	196.18
CI MTD	5.56	5.90	5.05	4.86	5.03	5.03	5.01	5.01
MSE minED	8.60	8.45	8.94	9.17	8.59	8.47	8.94	9.21
CI minED	1.89	1.85	1.91	1.90	1.92	1.92	1.95	1.95

Table C.16: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), mean squared error for the MTD (MSE MTD) and minimum effective dose (MSE minED), and median width of the 95% confidence intervals for the MTD (CI MTD) and the minimum effective dose (CI minED) in **Scenario III** for different settings of **SLOD based on the bivariate model**.

C.3 Scenario IV

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
0.6	7.85	0.01	0.01	0.00	0.00
1.2	9.71	0.08	0.10	0.03	0.04
2.0	12.15	0.58	0.66	0.28	0.55
3.0	14.37	3.08	3.23	1.26	1.32
4.0	16.43	10.66	11.44	5.18	5.07
5.3	16.30	26.45	27.11	18.68	17.69
7.0	**12.16**	**33.37**	**33.80**	**38.64**	**37.17**
9.3	3.85	20.46	20.24	30.35	29.95
12.4	0.29	4.78	2.96	5.41	6.97
16.5	0.00	0.53	0.44	0.17	1.23
22.0	0.00	0.00	0.00	0.01	0.01
29.4	0.00	0.00	0.00	0.00	0.01
none	6.90	0.00	0.00	0.00	0.00

Table C.17: Percentage of each dose being estimated as the *MTD* in **Scenario IV** for the **3+3 design** and different settings in **Bayesian ADEPT**; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	5.37	6.20	5.41	5.28	5.42	6.15	5.07	5.12
1.2	6.33	6.18	6.85	7.05	6.26	6.13	7.13	7.18
2.0	6.88	6.89	8.37	8.96	7.17	7.47	8.24	8.90
3.0	10.55	10.75	13.97	15.05	11.05	11.30	13.84	15.11
4.0	12.97	12.81	19.38	20.47	13.14	13.36	19.66	20.59
5.3	18.27	18.05	20.71	20.87	17.25	17.80	20.62	21.05
7.0	**18.66**	**18.75**	**13.34**	**13.34**	**19.27**	**18.09**	**13.56**	**13.42**
9.3	7.97	8.57	4.57	1.91	7.46	8.23	4.67	1.82
12.4	3.59	2.41	0.58	0.21	3.47	2.08	0.53	0.16
16.5	1.41	0.72	0.02	0.01	1.50	0.69	0.02	0.00
22.0	0.33	0.00	0.01	0.00	0.33	0.32	0.01	0.00
29.4	0.93	0.00	0.01	0.00	1.03	1.74	0.01	0.00
none	6.75	6.85	6.80	6.85	6.65	6.64	6.65	6.64

Table C.18: Percentage of each dose being estimated as the MTD in **Scenario IV** for **SLOD based on the logistic model** with different optimality criteria, design regions, and cohort sizes n; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	7.32	6.92	7.15	6.99	6.91	6.85	6.82	6.76
1.2	6.58	6.60	6.24	6.62	6.44	6.97	6.42	7.03
2.0	5.95	6.16	6.64	6.90	5.87	6.22	6.24	6.70
3.0	5.58	6.09	6.92	7.36	5.67	6.12	6.77	6.80
4.0	9.43	8.74	10.14	9.04	8.80	8.47	10.51	9.04
5.3	15.64	15.47	16.10	15.66	15.36	14.79	16.27	15.98
7.0	**20.73**	**19.48**	**20.00**	**19.97**	**20.88**	**19.49**	**20.41**	**20.18**
9.3	13.94	11.80	13.17	12.90	14.24	11.75	12.91	13.51
12.4	4.11	5.25	4.19	4.49	4.41	5.34	3.98	4.50
16.5	1.70	2.36	1.19	1.19	1.74	2.39	1.24	1.20
22.0	0.74	1.40	0.51	0.74	1.00	1.48	0.64	0.64
29.4	2.15	3.23	1.17	1.36	2.11	3.54	1.24	1.35
none	6.13	6.50	6.58	6.78	6.57	6.59	6.55	6.31

Table C.19: Percentage of each dose being estimated as the MTD in **Scenario IV** for **SLOD based on the 4 category proportional odds model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	5.44	6.38	5.41	4.92	5.43	5.89	5.30	5.36
1.2	5.21	5.37	5.78	6.01	5.03	5.18	5.64	5.88
2.0	4.35	5.26	6.11	7.38	4.59	5.85	6.15	6.71
3.0	5.23	4.92	9.77	10.58	5.43	5.89	9.91	10.76
4.0	9.18	9.49	18.82	19.40	10.02	10.51	19.60	19.92
5.3	21.01	21.92	22.90	25.02	21.16	22.19	22.99	25.33
7.0	**24.63**	**22.62**	**17.46**	**16.73**	**23.18**	**21.19**	**17.31**	**16.43**
9.3	12.86	12.22	6.14	2.71	13.20	11.81	5.68	2.53
12.4	3.15	2.67	0.91	0.21	3.10	2.52	0.74	0.21
16.5	1.40	0.66	0.06	0.01	1.35	0.65	0.05	0.01
22.0	0.15	0.21	0.01	7.03	0.21	0.32	0.01	6.86
29.4	0.68	1.31	0.01	4.92	0.67	1.41	0.01	5.36
none	6.71	6.97	6.62	6.01	6.63	6.59	6.61	5.88

Table C.20: Percentage of each dose being estimated as the MTD in **Scenario IV** for **SLOD based on the bivariate model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
$\overline{N}$	22.70	23.05	24.05	26.55	28.17
$\overline{N_{DLT}}$	3.52	7.26	7.57	8.03	8.44
$\overline{N_{>MTD}}$	1.13	7.20	6.84	8.04	7.63
MSE	15.10	5.35	4.88	4.64	6.12

Table C.21: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), and mean squared error (MSE) in **Scenario IV** for the **3+3 design** and different settings in **Bayesian ADEPT**.

	D-criterion				**c-criterion**			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	20.62	21.03	20.58	20.89	20.61	21.00	20.62	20.91
$\overline{N_{DLT}}$	3.54	3.18	3.33	3.37	3.35	3.10	3.33	3.37
$\overline{N_{>MTD}}$	1.88	0.98	1.47	1.39	1.55	0.78	1.46	1.38
MSE	18.87	22.00	12.54	12.67	19.52	22.59	12.42	12.60
CI	2.98	3.07	2.71	2.72	2.80	2.80	2.71	2.71

Table C.22: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of the 95% confidence interval for the MTD (CI) in **Scenario IV** for different settings of **SLOD based on the logistic model**.

	D-criterion				**c-criterion**			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	21.38	21.70	21.34	21.59	21.37	21.62	21.37	21.66
$\overline{N_{DLT}}$	4.82	4.36	4.57	4.65	4.77	4.31	4.64	4.70
$\overline{N_{>MTD}}$	2.79	1.89	2.42	2.39	2.75	1.83	2.52	2.48
MSE	26.45	34.71	20.54	22.33	26.83	36.74	21.02	21.88
CI	3.17	3.53	3.07	3.10	3.10	3.10	3.11	3.11

Table C.23: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of the 95% confidence interval for the MTD (CI) in **Scenario IV** for different settings of **SLOD based on the 4 category proportional odds model**.

	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	22.08	22.30	22.08	22.28	22.08	22.40	22.08	22.30
$\overline{N_{DLT}}$	4.22	3.62	3.70	3.72	4.02	3.54	3.72	3.73
$\overline{N_{>MTD}}$	2.93	1.60	1.76	1.58	2.54	1.38	1.81	1.63
MSE MTD	14.96	18.36	11.06	11.03	15.11	19.15	10.98	11.06
CI MTD	3.29	3.66	2.78	2.81	3.11	3.11	2.78	2.78
MSE minED	4.77	4.89	5.12	5.21	4.82	4.93	5.08	5.23
CI minED	2.54	2.51	2.79	2.71	2.52	2.52	2.84	2.84

Table C.24: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error for the MTD (MSE MTD) and minimum effective dose (MSE minED), and median width of the 95% confidence intervals for the MTD (CI MTD) and the minimum effective dose (CI minED) in **Scenario IV** for different settings of **SLOD based on the bivariate model**.

C.4 Scenario V

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
0.6	4.71	0.01	0.00	0.01	0.00
1.2	8.00	0.01	0.01	0.01	0.01
2.0	11.64	0.86	0.13	0.15	0.20
3.0	13.41	0.34	0.23	0.37	0.35
4.0	14.37	0.90	0.75	0.98	1.02
5.3	13.63	2.53	2.08	2.78	2.35
7.0	12.33	5.57	5.22	6.37	5.60
9.3	9.27	11.22	11.59	12.39	11.01
12.4	5.80	18.09	18.33	18.99	17.94
16.5	**2.90**	**23.68**	**24.14**	**23.09**	**24.05**
22.0	1.27	20.98	20.85	21.27	21.62
29.4	0.59	15.82	16.67	13.59	15.86
none	2.07	0.00	0.00	0.00	0.00

Table C.25: Percentage of each dose being estimated as the MTD in **Scenario V** for the **3+3 design** and different settings in **Bayesian ADEPT**; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	3.44	3.74	3.23	3.27	3.40	3.82	3.21	3.28
1.2	4.42	4.46	4.54	4.83	4.34	4.41	4.50	4.83
2.0	4.72	5.00	6.19	6.79	5.01	5.64	6.19	6.77
3.0	4.38	4.65	8.92	9.78	5.35	6.17	8.84	9.74
4.0	6.01	7.67	14.07	17.75	7.95	9.43	14.08	17.71
5.3	9.80	15.23	18.99	23.34	10.99	15.04	19.08	23.40
7.0	16.07	22.87	19.69	20.30	15.57	22.97	19.71	20.43
9.3	13.23	18.79	12.36	5.80	12.36	16.24	12.37	5.72
12.4	13.01	8.04	4.92	2.17	12.70	6.95	4.93	2.16
16.5	**7.76**	**2.17**	**1.77**	**2.27**	**7.15**	**2.03**	**1.77**	**2.30**
22.0	5.10	1.35	0.39	0.46	4.00	1.35	0.39	0.47
29.4	10.07	4.01	2.86	1.21	9.12	3.94	2.87	1.15
none	2.00	2.03	2.06	2.03	2.06	2.01	2.06	2.07

Table C.26: Percentage of each dose being estimated as the MTD in **Scenario V** for **SLOD based on the logistic model** with different optimality criteria, design regions, and cohort sizes n; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	4.74	4.92	4.74	4.77	4.74	4.82	4.74	4.95
1.2	7.95	7.60	7.92	7.81	7.95	7.85	7.94	7.90
2.0	9.81	9.94	9.91	9.92	9.87	9.91	9.90	9.76
3.0	10.42	10.79	10.75	10.67	10.45	10.79	10.70	11.03
4.0	9.93	10.07	10.89	10.74	10.06	10.08	10.90	10.67
5.3	9.79	10.31	10.58	10.53	9.82	10.27	10.53	10.39
7.0	9.61	10.37	10.49	10.12	9.41	10.34	10.46	9.95
9.3	8.81	10.51	8.70	9.36	8.73	10.06	8.77	9.28
12.4	7.32	8.25	7.08	7.50	7.45	8.28	7.09	7.61
16.5	**6.15**	**5.54**	**5.47**	**5.29**	**6.10**	**5.58**	**5.50**	**5.48**
22.0	4.71	3.41	4.11	4.46	4.69	3.48	4.16	4.36
29.4	8.71	6.27	7.31	6.78	8.68	6.52	7.26	6.56
none	2.05	2.02	2.05	2.05	2.05	2.02	2.05	2.06

Table C.27: Percentage of each dose being estimated as the MTD in **Scenario V** for **SLOD based on the 4 category proportional odds model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	3.69	4.15	3.62	3.74	3.68	4.06	3.58	3.89
1.2	5.24	4.97	5.29	5.48	5.19	4.57	5.29	5.40
2.0	4.98	5.78	5.78	6.44	5.05	6.14	5.50	6.48
3.0	4.53	4.76	7.10	6.90	4.93	4.36	7.27	7.56
4.0	3.73	3.58	9.73	10.92	4.45	5.20	10.30	10.58
5.3	3.73	5.92	14.61	16.60	5.11	7.72	14.45	17.61
7.0	6.81	13.66	17.16	20.95	7.93	14.26	17.40	20.61
9.3	11.31	19.84	15.72	17.97	11.35	19.52	15.44	17.80
12.4	15.75	19.96	12.34	5.20	15.34	18.94	12.19	4.88
16.5	**12.92**	**8.97**	**3.37**	**1.61**	**11.58**	**7.26**	**3.25**	**1.21**
22.0	10.44	1.93	1.45	0.48	9.73	1.72	1.46	0.35
29.4	14.58	4.45	1.61	1.61	13.43	4.06	1.65	1.60
none	2.29	2.03	2.22	2.10	2.23	2.19	2.22	2.03

Table C.28: Percentage of each dose being estimated as the MTD in **Scenario V** for **SLOD based on the bivariate model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
$\overline{N}$	26.37	40.63	42.45	39.30	41.74
$\overline{N_{DLT}}$	3.58	11.80	12.36	11.84	12.47
$\overline{N_{>MTD}}$	0.08	20.75	22.36	17.65	18.61
MSE	140.95	53.58	52.29	50.47	52.28

Table C.29: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), and mean squared error (MSE) in **Scenario V** for the **3+3 design** and different settings in **Bayesian ADEPT**.

	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	25.88	26.17	25.87	26.18	25.87	26.16	25.87	26.16
$\overline{N_{DLT}}$	3.16	3.00	3.24	3.26	3.09	2.96	3.25	3.26
$\overline{N_{>MTD}}$	0.44	0.03	0.95	1.03	0.28	0.03	0.95	1.02
MSE	102.87	111.55	126.99	135.92	106.80	116.84	126.83	135.85
CI	5.69	6.82	3.28	3.72	4.74	4.74	3.29	3.29

Table C.30: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of 95% confidence interval for the MTD (CI) in **Scenario V** for different settings of **SLOD based on the logistic model**.

	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	24.66	24.90	24.66	24.87	24.66	24.85	24.66	24.80
$\overline{N_{DLT}}$	4.53	4.46	4.54	4.56	4.53	4.47	4.54	4.56
$\overline{N_{>MTD}}$	0.60	0.14	0.77	0.70	0.60	0.14	0.77	0.70
MSE	131.55	130.25	133.03	131.70	131.72	130.67	132.86	131.85
CI	5.47	5.85	4.39	4.50	5.29	5.29	4.38	4.38

Table C.31: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of 95% confidence interval for the MTD (CI) in **Scenario V** for different settings of **SLOD based on the 4 category proportional odds model**.

	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	25.30	25.72	25.30	25.70	25.30	25.77	25.30	25.63
$\overline{N_{DLT}}$	3.79	3.51	3.59	3.61	3.71	3.45	3.58	3.61
$\overline{N_{>MTD}}$	1.73	0.10	1.20	1.07	1.40	0.07	1.21	1.07
MSE MTD	95.28	92.22	112.01	121.34	97.74	95.61	112.45	123.01
CI MTD	5.95	6.15	3.19	3.32	5.21	5.21	3.18	3.18
MSE minED	42.09	28.39	37.63	36.84	39.20	28.0	37.68	36.44
CI minED	9.07	8.32	7.95	7.54	8.49	8.49	8.28	8.28

Table C.32: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error for the MTD (MSE MTD) and minimum effective dose (MSE minED), and median width of 95% confidence intervals for the MTD (CI MTD) and the minimum effective dose (CI minED) in **Scenario V** for different settings of **SLOD based on the bivariate model**.

C.5 Scenario VI

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
0.6	0.04	0.00	0.00	0.00	0.00
1.2	0.26	0.00	0.00	0.00	0.00
2.0	0.82	0.02	0.01	0.00	0.00
3.0	1.99	0.00	0.00	0.00	0.00
4.0	4.30	0.03	0.03	0.00	0.02
5.3	8.69	0.27	0.24	0.09	0.11
7.0	15.75	1.63	1.71	0.97	0.89
9.3	23.40	8.20	8.81	6.40	5.76
12.4	24.01	23.87	22.57	23.18	20.42
16.5	**14.56**	**41.92**	**42.56**	**41.94**	**44.45**
22.0	5.13	22.46	22.29	25.26	25.59
29.4	1.03	1.61	1.79	2.16	2.76
none	0.01	0.00	0.00	0.00	0.00

Table C.33: Percentage of each dose being estimated as the MTD in **Scenario VI** for the **3+3 design** and different settings in **Bayesian ADEPT**; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	0.04	0.06	0.04	0.04	0.04	0.05	0.04	0.05
1.2	0.24	0.27	0.24	0.28	0.24	0.23	0.24	0.27
2.0	0.78	0.74	0.78	0.74	0.78	0.74	0.78	0.74
3.0	1.54	1.55	1.58	1.61	1.54	1.56	1.63	1.68
4.0	2.89	2.89	3.57	4.99	2.90	2.86	3.66	5.18
5.3	4.61	4.58	7.06	9.17	4.74	4.86	7.06	8.95
7.0	6.60	6.92	14.10	15.49	7.96	8.24	14.03	15.58
9.3	12.97	14.45	25.86	25.91	15.48	17.97	25.86	25.75
12.4	22.31	27.55	24.86	23.95	22.44	26.13	24.78	24.02
16.5	**25.94**	**24.31**	**15.89**	**14.82**	**24.20**	**23.17**	**15.89**	**14.81**
22.0	13.66	12.82	4.02	2.60	12.09	10.73	4.02	2.59
29.4	8.42	3.87	2.01	0.40	7.61	3.46	2.01	0.36
none	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Table C.34: Percentage of each dose being estimated as the MTD in **Scenario VI** for **SLOD based on the logistic model** with different optimality criteria, design regions, and cohort sizes n; 100 000 simulation runs.

Dose	D-criterion				c-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	0.04	0.07	0.07	0.06	0.06	0.05	0.06	0.04
1.2	0.22	0.27	0.18	0.21	0.23	0.27	0.16	0.16
2.0	0.82	0.80	0.88	0.82	0.79	0.85	0.86	0.91
3.0	1.72	1.88	1.62	1.75	1.70	1.90	1.72	1.87
4.0	3.62	3.78	3.62	3.25	3.56	3.47	3.49	3.36
5.3	6.29	6.59	6.26	7.02	6.32	6.53	6.28	6.93
7.0	8.91	9.49	10.40	10.78	8.79	9.60	10.39	10.67
9.3	14.07	14.37	16.87	17.50	13.03	12.61	16.71	16.76
12.4	21.37	22.06	23.81	25.07	19.30	20.79	23.33	23.68
16.5	**22.34**	**23.04**	**20.55**	**20.09**	**22.77**	**23.90**	**20.88**	**20.90**
22.0	14.67	12.91	11.08	9.72	17.19	15.07	11.42	10.91
29.4	5.93	4.74	4.66	3.73	6.26	4.96	4.70	3.81

Table C.35: Percentage of each dose being estimated as the MTD in **Scenario VI** for **SLOD based on the 4 category proportional odds model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
0.6	0.01	0.06	0.01	0.04	0.01	0.03	0.01	0.04
1.2	0.22	0.23	0.22	0.25	0.22	0.26	0.22	0.21
2.0	0.68	0.86	0.68	0.75	0.68	0.74	0.68	0.74
3.0	1.60	1.80	1.60	1.63	1.60	1.46	1.60	1.54
4.0	3.01	2.72	3.04	2.85	3.01	2.73	3.05	2.71
5.3	4.64	4.45	5.45	6.88	4.63	4.65	5.57	6.84
7.0	6.59	6.00	12.41	14.68	6.72	6.60	12.85	15.08
9.3	6.92	6.88	22.13	23.14	8.02	8.68	22.20	24.32
12.4	14.16	16.82	28.08	27.11	16.07	19.57	27.94	26.85
16.5	**32.41**	**33.19**	**18.97**	**16.70**	**31.33**	**31.03**	**18.60**	**15.74**
22.0	22.97	19.50	5.36	5.09	20.91	17.35	5.24	5.14
29.4	6.79	7.48	2.05	0.86	6.80	6.89	2.04	0.78
none	0.00	0.01	0.00	0.02	0.00	0.01	0.00	0.01

Table C.36: Percentage of each dose being estimated as the MTD in **Scenario VI** for **SLOD based on the bivariate model** with different optimality criteria, design regions, and cohort sizes n; 10 000 simulation runs.

Dose	3+3 Design	Bayesian ADEPT			
		variance gain function		patient gain function	
		n=1	n=2	n=1	n=2
$\overline{N}$	34.51	25.60	27.19	30.01	31.37
$\overline{N_{DLT}}$	3.53	7.50	7.96	9.07	9.40
$\overline{N_{>MTD}}$	0.29	10.94	12.39	12.11	12.40
MSE	57.69	19.66	19.99	19.44	19.70

Table C.37: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), and mean squared error (MSE) in **Scenario VI** for the **3+3 design** and different settings in **Bayesian ADEPT**.

	D-criterion				**c-criterion**			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	35.12	35.44	35.12	35.45	35.12	35.45	35.12	35.44
$\overline{N_{DLT}}$	3.64	3.26	3.53	3.60	3.51	3.16	3.53	3.60
$\overline{N_{>MTD}}$	2.84	1.56	2.42	2.63	2.47	1.27	2.42	2.62
MSE	50.08	44.26	54.79	57.80	51.01	45.98	54.96	57.92
CI	2.35	2.45	2.01	2.00	2.18	2.18	2.01	2.01

Table C.38: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of 95% confidence interval for the MTD (CI) in **Scenario VI** for different settings of **SLOD based on the logistic model**.

	D-criterion				**c-criterion**			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	33.83	34.07	33.91	34.27	33.84	34.16	33.91	34.29
$\overline{N_{DLT}}$	4.78	4.71	4.62	4.71	4.90	4.82	4.66	4.75
$\overline{N_{>MTD}}$	2.88	2.44	2.32	2.23	3.33	2.87	2.44	2.46
MSE	52.42	51.80	52.36	51.84	52.63	51.30	52.24	51.90
CI	2.03	2.04	2.01	2.01	2.06	2.06	2.01	2.01

Table C.39: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error (MSE), and median width of 95% confidence interval for the MTD (CI) in **Scenario VI** for different settings of **SLOD based on the 4 category proportional odds model**.

	D-criterion				L-criterion			
	des. region 1		des. region 2		des. region 1		des. region 2	
	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
$\overline{N}$	34.74	35.17	34.74	35.16	34.74	35.20	34.74	35.23
$\overline{N_{DLT}}$	4.96	4.54	4.38	4.50	4.85	4.40	4.37	4.52
$\overline{N_{>MTD}}$	5.41	4.11	3.47	3.63	5.07	3.63	3.48	3.63
MSE MTD	45.62	45.83	49.18	51.42	46.03	45.55	49.71	51.68
CI MTD	2.22	2.33	1.88	1.90	2.17	2.17	1.88	1.88
MSE minED	39.58	32.68	29.38	28.63	37.69	29.74	29.24	29.34
CI minED	3.05	3.19	2.97	2.96	2.98	2.98	2.97	2.97

Table C.40: Average number of subjects ($\overline{N}$), of observed DLTs ($\overline{N_{DLT}}$), of subjects treated with doses above the MTD ($\overline{N_{>MTD}}$), mean squared error for the MTD (MSE MTD) and minimum effective dose (MSE minED), and median width of 95% confidence intervals for the MTD (CI MTD) and the minimum effective dose (CI minED) in **Scenario VI** for different settings of **SLOD based on the bivariate model**.

Bibliography

Agresti, A. (1990). *Categorical Data Analysis*. Wiley series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

Atkinson, A. C. and Donev, A. N. (1996). *Optimum Experimental Designs*. Oxford Statistical Science Series. Oxford University Press, Oxford.

Cox, D. and Hinkley, D. (2000). *Theoretical Statistics*. Chapman & Hall / CRC, Boca Raton.

Dette, H. (1996). Lower bounds for efficiencies with applications. In Brunner, E. and Denker, M., editors, *Research Developments in Probability and Statistics: Festschrift zum 65sten Geburtstag von M.L. Puri*, pages 111–124. VSP Utrecht, The Netherlands.

Dragalin, V. and Fedorov, V. (2006). Adaptive designs for dose-finding based on efficacy-toxicity response. *Journal of Statistical Planning and Inference*, 136(6):1800–1823.

Dragalin, V., Fedorov, V. V., and Wu, Y. (2006). Optimal designs for bivariate probit model. Technical report, GlaxoSmithKline Pharmaceuticals, Collegeville, PA.

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

Fedorov, V. V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*, volume 125 of *Lecture Notes in Statistics*. Springer, New York.

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

Gerke, O. and Siedentop, H. (2007). Optimal phase I dose escalation trial designs in oncology - A simulation study. *Statistics in Medicine*, 27:5329–5344.

Habermann, S. J. (1980). Discussion of Dr. McCullagh's paper. *Journal of the Royal Statistical Society, Series B*, 42:136–137.

Hardy, G., Littlewood, J., and Pólya, G. (1988). *Inequalities*. Cambridge University Press, Cambridge.

Hosmer, D. W. and Lemeshov, S. (1989). *Applied Logistic Regression*. Wiley series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

Ivanova, A. (2006a). Dose-finding in oncology – nonparametric methods. In Ting (2006a), pages 49–58.

Ivanova, A. (2006b). Escalation, group and A + B designs for dose-finding trials. *Statistics in Medicine*, 25:3668–3678.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society, Series B*, 21:272–319.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 42:363–366.

Kotz, S., Balakrishnan, N., and Johnson, N. (2000). *Continuous Multivariate Distributions, vol. 1.* Wiley, New York.

Lin, Y. and Shih, W. J. (2001). Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trial. *Biostatistics*, 2:203–215.

Liu, G. (2006). *Sequential Designs for Logistic Phase I Clinical Trials*. PhD thesis, University of Maryland Graduate School, Baltimore, MD.

Liu, G., Rosenberger, W. F., and Haines, L. M. (2009). Sequential designs for ordinal phase I clinical trials. *Biometrical Journal*, 51:335–347.

MacDougall, J. (2006). Analysis of dose-response-studies – e_{max} model. In Ting (2006a), pages 127–145.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, 42:109–127.

Modi, M. (2006). Dose finding studies in phase i and estimation of maximally tolerated dose. In Ting (2006a), pages 30–48.

National Cancer Institute (2006). *Common Terminology Criteria for Adverse Events*.

O'Quigley, J., Pepe, M., and Fisher, L. (1990). Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics*, 46:33–48.

Pázman, A. (1986). *Foundations of Optimum Experimental Design*. Mathematics and its Applications. D. Reidel Publishing Company, Dordrecht.

Perevozskaya, I., Rosenberger, W. F., and Haines, L. M. (2003). Optimal design for the proportional odds model. *The Canadian Journal of Statistics*, 31(2):225–235.

Pratt, J. W. (1981). Concavity of the log likelihood. *Journal of the American Statistical Association*, 76:103–106.

Rabie, H. and Flournoy, N. (2004). Optimal designs for contingent response models. In Bucchianico, A., Läuter, H., and Wynn, H., editors, *MoDa 7 Advances in Model-Oriented Design and Analysis*, pages 133–142.

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

Silvapulle, M. J. (1981). On the existence of maximum likelihood estimators for the binomial response model. *Journal of the Royal Statistical Society, Series B*, 43:310–313.

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

Thall, P. F. and Cook, J. D. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*, 60:684–693.

Tighiouart, M. and Rogatko, A. (2006). Dose-finding in oncology – parametric methods. In Ting (2006a), pages 59–72.

Ting, N., editor (2006a). *Dose Finding in Drug Development*, Statistics for Biology and Health, New York. Springer.

Ting, N. (2006b). Introduction and new drug development process. In Ting (2006a), pages 1–17.

White, L. V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika*, 60:345–348.

Whitehead, J. and Williamson, D. (1998). Bayesian decision procedures based on logistic regression models for dose-finding studies. *Journal of Biopharmaceutical Statistics*, 8:445–467.

Witting, H. (1985). *Mathematische Statistik I*. B. G. Teubner, Stuttgart.

Zhou, Y. and Whitehead, J. (2002). *Bayesian ADEPT: Operating Manual*. Medical and Pharmaceutical Statistics Research Unit, University of Reading, U.K.