





This is to certify that the dissertation entitled

ADAPTIVE DESIGNS WITH COVARIATES

presented by

GEORGE SIRBU

has been accepted towards fulfillment of the requirements for the

Ph.D.	degree in Statistics and Probability	
	(722)	
	Major Professor's Signature	
	7-15-04	
	Date	

MSU is an Affirmative Action/Equal Opportunity Institution

LIBRARY Michigan State University

PLACE IN RETURN BOX to remove this checkout from your record. TO AVOID FINES return on or before date due. MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

6/01 c:/CIRC/DateDue.p65-p.15

ADAPTIVE DESIGNS WITH COVARIATES

By

George Sirbu

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2004

ABSTRACT

ADAPTIVE DESIGNS WITH COVARIATES

by

George Sirbu

The potential benefits of adaptive allocation have been recognized as a great contribution especially for clinical trials. Implementation of these allocation schemes eases the ethical problem involved in trials on human subjects. While responseadaptive randomization does not eliminate the ethical problem of randomizing patients to the inferior treatment, it can mitigate it by making the probability of assignment to the inferior treatment smaller. The relatively new techniques of response-adaptive randomization are attractive to industry and also to the Food and Drug Administration. An essential feature is to balance the conflict between information gathering - collective ethics - and immediate payoff - individual ethics. In this sense response-adaptive randomization represents a middle ground between community benefit and individual patient benefit.

In this dissertation we concentrate on designs with covariate information and response adaptive randomization procedures. In some studies it is desired to conduct an analysis that is 'adjusted' for other covariates. Stratification is one approach to adjust for covariates or to increase the efficiency by accounting for a highly influential covariate. This approach however, is only applicable to qualitative covariates, or discretized quantitative covariates, and they must be few in number or else the number of strata grows exponentially. In many respects it is more natural to perform an adjustment using a regression model that allows for both qualitative and quantitative covariates simultaneously.

The problem consists of choosing in a sequential manner one of two treatments while we continuously observe the information from the process - the response and the covariates. We try to choose the best allocation while at the same time we learn from the process. We are able to prove strong consistency and asymptotic normality for maximum quasi-likelihood estimators of regression parameters in generalized linear models with a condition of smoothness for the link function. The key for the applicability of the results is to have a martingale difference structure for the errors of the model and to have a matrix for the covariates that is compact enough, where the compactness condition is expressed in terms of the eigenvalues of the design matrix.

The results are general enough to allow us to define the design in a large variety of situations. For obvious ethical reasons, we investigate also what number of patients allocated to the inferior treatment we can expect from the design under some given underlined distributions for the covariate. Various Monte Carlo simulations are used to evaluate the performance of the design for suitable choices of the parameters.

To my dear wife Mihaela.

ACKNOWLEDGEMENT

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I am deeply indebted to my supervisors Dr. C. Page and V. Melfi whose help, stimulating suggestions and encouragement helped me in all the time of research for and writing of this thesis. I wish to thank the members of my committee Dr. H. Salehi and R. Erickson, who have given me valuable suggestions.

It is impossible to have my research career without my parents' love and support, as well as my colleagues and friends' encouragement. I appreciate all their friendships and their collective encouragement to finish this dissertation.

Especially, I would like to give my special thanks to my wife Mihaela whose patient love enabled me to overcome the obstacles through this journey.

Table of contents

List of Tables vii			vii	
Li	List of Figures vii			
1	Cha 1.1 1.2 1.3	Apter 1 Introduction	1 1 3 11	
2	Cha 2.1 2.2 2.3	apter 2 Introduction	14 14 15 19 21	
3	Cha 3.1 3.2 3.3 3.4	Apter 3 Introduction A Generalized Linear Model Relating the Response, the Adaptive Design and the Covariates A Consistency Result An Asymptotic Normality Result	27 27 29 35 40	
4	Cha 4.1 4.2	apter 4 Introduction A Normal Response Example 4.2.1 An Evaluation Function for the Number of Patients on the Inferior Treatment 4.2.2 Simulations of the Design Conclusions and Future Work	43 43 48 51 60 81	
Aj	Appendix			
Bi	Bibliography			

List of Tables

4.1	Allocation probabilities to treatment A	62
4.2	Simulation results for $\hat{eta}_1, \hat{eta}_2, \hat{eta}_3, \hat{eta}_4,$ in the case $n=30$	65
4.3	Simulation results for $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$, in the case $n = 50 \ldots \ldots$	66
4.4	Simulation results for $\hat{eta}_1, \hat{eta}_2, \hat{eta}_3, \hat{eta}_4,$ in the case $n=100$	67
4.5	Simulation results for C_{30} , in the case $n=30$	68
4.6	Simulation results for C_{50} , in the case $n = 50 \ldots \ldots \ldots \ldots$	69
4.7	Simulation results for C_{100} , in the case $n = 100 \dots \dots \dots \dots$	70
4.8	Simulation results for C_{50} , in the case $n = 50 \ldots \ldots \ldots \ldots$	77
4.9	Simulation results for $C_{50}/40$, in the case $n=50$	78
4.10	Simulation results for C_{100} , in the case $n = 100 \ldots \ldots \ldots \ldots$	79
4.11	Simulation results for $C_{100}/90$, in the case $n = 100 \dots \dots \dots \dots$	80

List of Figures

4.1	Graph of function $f_1(y) = \frac{\exp(y)}{1 + \exp(y)}$	51
4.2	Illustration of Corollary 4.2.2	59
4.3	The regression lines for treatment A and B	61
4.4	A simulation design	63
4.5	Histogram of 10000 replications for bias of $\hat{\beta}_1$ when $n = 100, \sigma = 0.1, 0.5, 1, 2$	71
4.6	Histogram of 10000 replications for bias of $\hat{\beta}_2$ when $n = 100, \sigma = 0.1, 0.5, 1, 2$	72
4.7	Histogram of 10000 replications for bias of $\hat{\beta}_3$ when $n = 100$, $\sigma = 0.1, 0.5, 1, 2$	73
4.8	Histogram of 10000 replications for bias of $\hat{\beta}_4$ when $n = 100$, $\sigma = 0.1, 0.5, 1, 2$	74
4.9	Histogram of 10000 replications for estimates of C_{100} when $n = 100$, $\sigma = 0.1, 0.5, 1, 2$	75

Chapter 1

Literature Review

1.1 Introduction

The current dissertation is motivated by the following scenario: in the case of a clinical trial when two drugs, A and B, are to be evaluated, a design is wanted such that the treatment assignment takes in consideration possible covariates and past responses. A design that incorporates in the allocation rule the information obtained from past observations is called an *adaptive design*. We will review in this chapter some of the designs proposed in the literature and we will see later chapters how we can construct a design for the problem formulated above.

As usually in the case of a design problem, we want to balance two goals. On one hand we want the design to be 'ethical' in the sense that the allocation should be done toward the treatment that shows the best performance thus far in the trial (this can be thought of as an 'individualistic goal' as defined by Sarkar(1991)). On the other hand we want the design to allow us to draw reliable statistical inference from the data to be collected (this can be thought of as an 'utilitarian goal' as also defined by Sarkar(1991)). Unfortunately the utilitarian and individualistic goals are usually conflicting. The adaptive designs may address both concerns arisen before. The adaptiveness of the design reflects that the treatment allocation is based on responses available thus far in the trial. But the problem is more complicated here because we set our goal to make the design dependent on covariates too. As one may guess, the covariate information should make the design more '*ethical*' but at the same time it is likely that it complicates the statistical inference.

In the current dissertation we will develop and study a new design that will satisfy the goals described before. In the second chapter we will look at some general properties of the distribution of responses in the case of an adaptive design with covariates. In chapter three we will prove some asymptotic results for such designs in the case of a specific model for the responses. In the final chapter we will look at some particular designs and how to evaluate them from an 'ethical' point of view and simulation results will be shown.

The dissertation work concentrates on designs where the response from the treatment depends on a vector of covariates. Such designs will henceforth be referred as *covariate designs*, as opposed to *non-covariate designs* which do not take into consideration covariate information on the patients involved in the clinical trial. The other focus of the dissertation is on the adaptiveness of the design as described before. Following Rosenberger's(2002) classification, the designs where the allocation is done based on the previous treatment allocations, used primarily when a balanced allocation is desired, are called *allocation-adaptive designs*. The designs where the allocation is done based on the previous responses from the treatments, primarily used when ethical considerations are preferred, are called *response-adaptive designs*.

Before presenting and studying the new design we begin by reviewing some of the procedures that have already been proposed in the literature. This review can be regarded as an introduction and motivation for the new procedure proposed in this dissertation; it will be particularly useful in showing what are the novel aspects of the new design. Also it introduces the designs that are to be used as basic points of reference for the new design.

1.2 A Short Review of Adaptive Designs

The main focus of this section is to review some of the adaptive designs available in the literature. Before we start to review some of these, is important to acknowledge the importance of the randomization in a clinical trial.

The randomized clinical trial guards against systematic bias. Another property of randomization is that it promotes comparability among the study groups. Such comparability can only be attempted in observational studies by adjusting for or matching on known covariates. Moreover, the act of randomization provides a probabilistic basis for an inference from the observed results.

In the remainder of this section, unless otherwise noted, it is assumed that patients arrive sequentially and that each patient can be assigned to exactly one of two treatments. Although Many of these designs can be modified to handle more than two treatments.

Complete randomization is a fair coin-tossing assignment: patients have equal probability to be assigned to either of the two treatments. Sometimes it is desired that the final allocation is exactly equal between the treatments. When the total sample size, n, is known, a truncated binomial design [Blackwell and Hodges(1957)] will achieve the goal of equal allocation. The design is simply to use complete randomization until one treatment has been assigned n/2 times; all subsequent patients will receive the opposite treatment. An alternative is the random allocation rule [Lachin(1988)]. Let $(t_k)_{k=1,...,n}$ be the sequence of random assignments (where $t_k = 1$ if the k^{th} patient receives treatment A and $t_k = 0$ otherwise) and we denote by N_k^A the number of patients randomized to treatment A after k patients have been allocated. The design is defined by the following:

$$P(t_1 = 1) = \frac{1}{2}$$

$$P(t_k = 1) = \frac{\frac{n}{2} - N_{k-1}^A}{n - k + 1}$$

The difference between the last two designs is that the treatment sequences will not be equiprobable for the truncated binomial design. These last two designs can be embedded in a larger class of designs called *restricted randomization designs*. In this case the treatment assignments $(t_k)_{k=1,...,n}$ are dependent with a variance-covariance matrix $\Sigma_t \neq I/4$. They are used when is desired to have equal numbers of patients assigned to each treatment group.

The biased coin design proposed by Efron(1971) is a modification of the complete randomization in the sense that it is a biased coin-tossing. Let $D_k = N_k^A - N_k^B$ be the difference between the number of patients who received treatment A and those who received treatment B after k patients have been randomized. Fix a constant p in the interval [0.5,1). Then the design is described by:

$$P(t_{k} = 1) = \begin{cases} 1 - p & \text{if } D_{k-1} > 0\\ 1/2 & \text{if } D_{k-1} = 0\\ p & \text{if } D_{k-1} < 0 \end{cases}$$

Wei(1978) noted that a disadvantage of this procedure is that, in assigning the next patient to a treatment, the allocation policy neither takes into consideration the number of patients treated thus far, nor does it discriminate between small and large absolute values of D_k . He proposed an *adaptive biased coin design* as follows: let $h: [-1, 1] \rightarrow [0, 1]$ be a non-increasing function such that h(x) = 1 - h(-x) for

all $x \in [-1, 1]$. Then

$$P(t_k = 1) = h(\frac{D_{k-1}}{k-1})$$
.

This allocation policy forces an extremely imbalanced experiment to be balanced but tends to complete randomization as the difference D_k tends to zero.

All previous designs have in common that they are not response-adaptive, i.e. they do not depend on the observed responses from the trial. Their goal is to randomize the treatment and some balance is required. The next designs that we'll present will be response-adaptive. The response-adaptive randomization can be used when various considerations make it desirable to have unequal numbers of patients assigned to treatments. For example, is very common in clinical trials to have more patients assigned to the superior treatment. Other examples are driven by optimality criteria, as they will be described below, which results in imbalanced allocation.

A well-known such design is the randomized play-the-winner rule (RPW) introduced by Wei and Durham(1978). The design can be described with an urn model as follows. The urn has initially u balls of two types corresponding to the two treatments. When a patient enters the study, a ball is randomly drawn from the urn and the treatment is assigned according to the ball type. The ball is returned to urn and the response is observed. If a success has occurred, add β balls of the initial ball type selected and α balls of the other type. If a failure has occurred, add α balls of the initial ball type selected and β balls of the other type. The rule is denoted by RPW(u, α, β). Different choices for the pair (α, β) give different levels of compromise between balance and allocation to the better treatment. A design used in practice is the RPW(1,0,1). Later many authors have modified and generalized the RPW to achieve less selection bias or less variability or some other optimization criteria. We mention here for example the *birth and death urn* by Ivanova, Rosenberger, Durham and Fluornoy(2000). A good source for a review of urn models is Rosenberger(2002).

So far the designs presented have intuitive motivation and can be completely non-parametric but are not derived from optimal considerations. Another approach for the response-adaptive design is based on optimal allocation targets, where a specific criterion is optimized based on a population response model.

The approach we take here is similar to Melfi, Page and Geraldes(2001). Let $\pi = \pi(\cdot, \cdot)$ be a function taking values in (0,1), and let θ^A and θ^B be two parameters for the distribution of the responses from treatment A, respectively B. Assume that there is an unknown but estimable optimal proportion $\pi = \pi(\theta^A, \theta^B)$ of the observations that should be allocated to treatment A. Let $\hat{\pi}_{k-1}$ be an estimate of π based on the first k - 1 observations. Then the rule is simply to allocate according to the estimate of the optimal proportion:

$$P(t_k=1)=\hat{\pi}_{k-1}$$

The optimal criterion can be defined in various ways. For example if the responses are normally distributed with standard deviations σ^A , σ^B and we want to minimize the variance of the difference of the sample means when the total sample size is fixed then the optimal target is the proportion $\pi = \frac{\sigma^A}{\sigma^A + \sigma^B}$ and thus the parameters to be estimated at each step are $\theta^A = \sigma^A$ and $\theta^B = \sigma^B$.

Another criterion is given by Jenison and Turnbull(2000). Let μ^A, μ^B be the mean responses from the treatment A, respectively B, and let $\hat{\mu}_n^A, \hat{\mu}_n^B$ be their estimators after *n* observations. Consider $\pi = N^A/n$, where $N_n^A + N_n^B = n$. The optimal π is found by minimizing the function $u(\mu^A, \mu^B)N^A + v(\mu^A, \mu^B)N^B$, with the variance of $\hat{\mu}_n^A - \hat{\mu}_n^B$ fixed. For example if the responses are dichotomous, so we can clearly define a success/failure, and we want to minimize the expected number

of treatment failures, then we choose $u(\mu^A, \mu^B) = 1 - \mu^A, v(\mu^A, \mu^B) = 1 - \mu^B$ and in this case we obtain $\frac{\pi}{1-\pi} = \sqrt{\frac{\mu^A}{\mu^B}}$.

Melfi and Page(2000) give an elegant method for proving consistency and asymptotic normality of estimators for a wide class of response-adaptive designs. We will see in Chapter 2 and Chapter 3 how some results can be generalized to the case where covariates are incorporated in the design.

Most of the work done in the area of the covariate adaptive designs involves the case when the covariates are categorical. Obviously continuous covariates can also be considered by grouping in an appropriate way the possible values of the covariates into finitely many levels. A general approach is to form strata with combinations of the categorical covariates and then apply the adaptive designs within each stratum. Both Efron(1971) and Wei(1978) mentioned in their papers the procedure to extend their designs to the covariate case.

Pocock and Simon(1975) suggest an allocation rule which can be viewed as a generalization of Efron's biased coin design to more than two treatments and several covariates. The design relies on a function G which measures the total amount of imbalance (in the distribution of the treatment numbers within strata). Treatments are then ranked according to their G-values. Their procedure is shown to enable treatments to be balanced across strata more effectively. But as Pocock and Simon(1975) note, a major difficulty with this approach is that number of strata increases rapidly as the number of covariates and their levels increase.

Geraldes(1999) proposes two new designs. The first one is the *covariate adaptive weighted differences design* which incorporates covariates and can be viewed as a generalization of the adaptive biased coin design of Wei(1978) by crossing-over information from responses of patients from stratum to stratum. The second one is the *covariate randomized play-the-winner rule* which corresponds to a multiple urn model, each urn representing a stratum. It allows the responses of patients in one stratum to change the composition of the urns corresponding to the other strata.

Instead of concerns about balancing treatment assignments across strata, one can take an entirely different approach and find an allocation rule that optimizes some criterion, such as minimizing the variance of the estimated treatment effect in the presence of covariates. Such a rule would necessarily require the specification of a model linking the covariates and the treatment effect.

As an example, Atkinson(1982) chooses a standard linear regression model:

$$E(Y_i) = x'_i\beta, \qquad i = 1, 2, \dots, n$$

where the Y_i are independent responses with $Var(\mathbf{Y}) = \sigma^2 \mathbf{I}$ and x_i includes a treatment indicator and selected covariates of interest. Then $Var(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$, where $\mathbf{X}'\mathbf{X}$ is the dispersion matrix. For the construction of the optimal design, we wish to find the *n* points of experimentation at which some function is optimized. The D_A -optimal design uses D_A -optimality that maximizes

$$D_A = |\mathbf{A}'\mathbf{M}^{-1}\mathbf{A}|^{-1}$$

where $\mathbf{M} = \mathbf{X}'\mathbf{X}/n$ and \mathbf{A} is a matrix of contrasts. He proposes then the following sequential design: assuming that n patients have been already allocated, the n + 1patient is assigned to the treatment that maximizes $D_A(\xi_n)$ evaluated at the n-point design ξ_n . He also compares the D_A criterion to *D*-optimality, which maximizes the log determinant of \mathbf{M} .

Clayton(1989) proposes a covariate model for a problem with the responses of the patients governed by Bernoulli distributions. The model can be formally described as follows. Let Y_k^A and Y_k^B denote the potential dichotomous responses of the k^{th} patient to treatment A and B, respectively. Before assigning patient k, a covariate

 X_k can be observed on the patient. It is assumed that the covariate random variables X_1, X_2, \ldots are i.i.d. with a known distribution F. He considered then a function H() for linking the responses to the covariates:

$$P(Y_k^A = 1 | X_k) = H(\alpha + \beta X_k)$$
$$P(Y_k^B = 1 | X_k) = H(c + dX_k)$$

where α and β are unknown constants and c and d are known constants. Following a Bayesian approach, he assumes that prior information regarding α and β is available, given by a probability distribution. The worth of an allocation(strategy) is defined as the expected sum of the first n observations for all possible histories resulting from the allocation policy. A strategy is called optimal if it yields the maximal worth. The research focuses on the determination of the structural characteristics of exactly optimal strategies. Yang and Zhu(2002) generalize the problem to the case when the responses are continuous and they propose a design so that they are able to prove that the strategy is strongly consistent in the sense that the accumulated reward is asymptotically equivalent to that based on the best treatment. Their approach employs nonparametric regression procedures for estimating the dependence of the rewards on the covariates for the treatments. It uses a randomized allocation scheme to control the trade-off between the tendency to use the currently most promising treatment and further exploration to find the treatment that is truly the best.

Lai and Wei(1982) consider the multiple regression model:

$$y_n = \beta_1 x_{n1} + \ldots + \beta_p x_{np} + \epsilon_n, \qquad n = 1, 2, \ldots$$

where the ϵ_n are unobservable random errors, β_1, \ldots, β_p are unknown parameters and y_n is the observed response corresponding to the design levels x_{n1}, \ldots, x_{np} . Let \mathbf{b}_n be the least squares estimators of $\beta = (\beta_1, \ldots, \beta_p)$. The statistical properties of the estimators \mathbf{b}_n are well understood in the case where the design levels x_{ij} are non-random constants, but there is a much less definitive theory for the case where the covariate vectors $\mathbf{x}_n = (x_{n1}, \ldots, x_{np})$ are sequentially determined random vectors. They consider an adaptive design in the sense that \mathbf{x}_n at each stage n depends on the previous observations $\mathbf{x}_1, y_1, \ldots, \mathbf{x}_{n-1}, y_{n-1}$. They give sufficient conditions to prove strong consistency and asymptotic normality for \mathbf{b}_n . They also give an interesting counterexample to show that their conditions are in some sense the weakest possible. Here is the counterexample:

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i$$

where $\epsilon_1, \epsilon_2, \ldots$ are i.i.d. random variables with $E\epsilon_1 = 0, E\epsilon_1^2 = 1$. The regressors are defined inductively by:

$$x_1 = 0, \qquad x_{n+1} = \overline{x}_n + c\overline{\epsilon}_n$$

where $c \neq 0$ is a real constant and $\overline{\epsilon}_n, \overline{x}_n$ denote the arithmetic means. Then it can be proved that the assumptions are violated and the least squares estimate b_{n2} of β_2 converges a.s. to $\beta_2 - c^{-1}$.

Chen, Hu and Ying(1999) consider a similar problem in the context of the generalized linear model. The key difference is obviously the nonlinearity in the link function. They handle this by establishing a local inverse function theorem. In doing so, they show that the minimum conditions of Lai and Wei(1982) in conjunction with an additional assumption, essentially to offset the nonlinearity, ensure strong consistency in their case, too.

We will consider in Chapter 3 a generalization of these last two designs. In their

case the adaptiveness refers to the dependence of the covariate vector \mathbf{x}_n on the observations $\mathbf{x}_1, y_1, \ldots, \mathbf{x}_{n-1}, y_{n-1}$. In our case we will consider the problem when the vector \mathbf{x}_n has two components: a treatment indicator t_n and a vector of patient covariates. Then the choice of the treatment, t_n , is made not only on the previous observations $\mathbf{x}_1, y_1, \ldots, \mathbf{x}_{n-1}, y_{n-1}$, but also on the values observed from the current patient covariate.

1.3 Applications

Medical investigators who wish to perform clinical trials currently have a wide variety of adaptive allocation procedures at their disposal. Still, very few clinical trials based on adaptive designs have been reported to the literature. We mention some of them here.

Cornell, Landenberger and Bartlett (1986) reported on an adaptive clinical trial to test the efficacy of extracorporeal membrane oxygenation (ECMO) for the treatment of persistent pulmonary hypertension of newborn infants. The design used was RPW(1,0,1). The trial was stopped after 12 total patients using a stopping rule with 1 patient allocated to the conventional therapy and the rest to the ECMO treatment. The subsequent analysis of the ECMO trial data generated controversy, the foremost question raised was whether two treatments can adequately be compared when only one patient was assigned to one of the treatments. Much of the criticism of adaptive designs has centered on this trial and this is unfortunate because this is exactly the type of trial where response-adaptive randomization would be particularly advantageous. It is known that the RPW(1,0,1) is highly variable, particularly when $p_A + p_B > 1.5$ which was the case for the ECMO trial, and the variance depends on the initial composition of the urn. In retrospect, starting with more than one ball of each type should have resulted in a more balanced trial. Another trial was the fluoxetine trial reported by Tamura, Faries, Andersen and Heiligenstein (1994). They used an RPW(1,0,1) design in a clinical trial of fluoxetine versus placebo for depressive disorder. The trial was stratified by normal and shortened rapid eye movement latency (REML), so two urns were used in randomization. In order to avoid the problem encountered in the ECMO trial, the first six patients in each stratum were assigned using a permuted block design. The trial was stopped after 61 patients had responded in accordance with surrogate criterion. The trial randomized a total of 89 patients: 21 fluoxetine patients and 20 placebo patients in the shortened REML stratum, 21 fluoxetine and 21 placebo patients in the normal REML stratum. The primary outcome was analyzed using Monte-Carlo randomization-based analysis.

Rosenberger(1999) discusses conditions under which the use of response-adaptive randomization is reasonable. We note some of them here:

- The treatments have been evaluated previously for toxicity. This is important to ensure that the response-adaptive randomization does not place more patients on a highly toxic treatment.
- Delay in response is moderate, allowing the adaptation to take place.
- Duration of the trial is limited and recruitment can take place during most or all of the trial.
- Modest gain in terms of treatment successes is desirable from an ethical standpoint.
- The experimental therapy is expected to have significant benefits to the public health.

The choice of a design should be driven by the simplicity of implementation but also by its statistical properties and the nature of the clinical trial. We hope that the new design proposed in this dissertation can be successfully used some future clinical trials.

Chapter 2

Adaptive Designs with Covariates: Model and Marginal Distributions

2.1 Introduction

In this chapter we introduce the problem of adaptive designs with covariates. We will look only at the response distributions without assuming any parametric model relating the response to the covariates and in this sense a wide class of examples is covered. Assume that we are in the case of clinical trials where the patients are randomly assigned, one at a time, to one of the two treatments according to a response-adaptive rule. We assume that we work under the following conditions:

- Patients arrive sequentially.
- Each patient can be assigned to either one of the two treatments and will be assigned to exactly one of them.
- The response from patients is observed immediately, prior to the assignment of the treatment to the next patient.

• The response depends on a vector of covariates. The covariates are observed prior to the treatment assignment.

Our goal is to show how a design for a response-adaptive allocation where the treatment allocation depends on the previous responses and the current covariates of the patient influences the distribution of the responses.

In contrast to the case when the adaptive design is without covariates, the new setup introduces a more complex dependence structure in the following sense: the response is dependent on the vector of the covariates and the treatment allocation and at the same time the treatment allocation is dependent on the previous responses. Of course that in the case when we ignore the covariates from the design, the first dependence that we described is not present.

2.2 The Model and Notation

In this section we introduce the notation for the model we will use through the rest of the chapter. As we described before, suppose that patients arrive sequentially and they are allocated to one of two treatments, say A or B. For each $k \ge 1$, define t_k to be 1 or 0 according to whether the k^{th} patient is assigned to treatment A, respectively treatment B:

$$t_{k} = \begin{cases} 1 & \text{if } k^{th} \text{ patient is assigned to treatment A} \\ 0 & \text{if } k^{th} \text{ patient is assigned to treatment B} . \end{cases}$$

Let Y_k^A, Y_k^B denote the potential response of the k^{th} patient for treatment A, respectively B. Recall from our initial assumptions that, for all $k \ge 1$, exactly one of the pair (Y_k^A, Y_k^B) is actually observed.

Let $X_k = (X_{k1}, \ldots, X_{kp})$ denotes the *p*-dimensional covariate vector of the k^{th}

patient. Our goal is to have an adaptive design where the responses can be dependent on covariates. Thus the distribution for Y_k^A, Y_k^B can be conditioned on the X_k . We consider then the following conditional distributions:

(2.1)
$$Y_k^A | X_k = x \sim F_x^A$$
$$Y_k^B | X_k = x \sim F_x^B .$$

Note that we do not assume or require any independence structure between Y_k^A and Y_k^B . In fact there may be a strong correlation between the responses observed from treatment A and B, since they are observed from the same patient.

In most cases the allocation is randomized. It is useful then to let $(U_k)_{k\geq 1}$ denote a sequence of i.i.d. random variables whose common distribution is uniform on (0, 1) and is independent of the other variables in the model (Y_k^A, Y_k^B, X_k) . Then the allocation is defined as:

$$(2.2) t_k = I[U_k < \hat{\pi}_k]$$

where the $I[\cdot]$ denotes the indicator function and $\hat{\pi}_k$ is some number. We will see later in chapter 4 how we can define $\hat{\pi}_k$ in a suitable way. We use the index k for $\hat{\pi}_k$ to emphasize that $\hat{\pi}_k$ is computed based on the covariate X_k , but we have to keep in mind that in fact $\hat{\pi}_k$ is computed based only on the information from the first k-1 patients and the current covariate X_k .

Assume now the existence of an increasing sequence of σ -algebras, $\{\mathcal{F}_k\}_{k\geq 1}$ such that the triplet $(Y_k^A, Y_k^B, X_k, t_k, U_k)$ is \mathcal{F}_k -measurable. We can define them as:

(2.3)
$$\mathcal{F}_k := \sigma(Y_i^A, Y_i^B, X_i, t_i, U_i : 1 \le i \le k) .$$

Think of the σ -algebra \mathcal{F}_k as all possible information available after k patients have

been allocated and their responses and covariates observed.

Consider the increasing sequence of σ -algebras $\{\mathcal{G}_k\}_{k\geq 1}$ defined as follows:

(2.4)
$$\mathcal{G}_k := \mathcal{F}_k \vee \sigma(X_{k+1}) \vee \sigma(U_{k+1}) .$$

We have to define a new σ -algebra so that the treatment indicator variable becomes measurable. In the view of our goal stated in the beginning of the chapter, think of the σ -algebra \mathcal{G}_k as the information to use for allocating the treatment for the $(k+1)^{th}$ patient. Therefore we will make use of \mathcal{G}_k as the σ -algebra with respect to which t_{k+1} becomes measurable. To understand better, observe what happens without adding $\sigma(X_{k+1})$ to the filtration $\mathcal{F}_k \vee \sigma(U_{k+1})$. If we require only that t_{k+1} is measurable with respect to $\mathcal{F}_k \vee \sigma(U_{k+1})$, as in the usual case of adaptive designs without covariates, then the treatment allocation depends only on the previous responses without taking in consideration the covariate value of the current patient. But this was not our proposed goal, so the definition of the new algebra plays a crucial role in following developments.

It is also useful to define the σ -algebra \mathcal{J}_k :

(2.5)
$$\mathcal{J}_k := \mathcal{F}_k \vee \sigma(X_{k+1}) .$$

Observe then from the definition of the sequence of the uniform variables $(U_k)_{k\geq 1}$ that U_{k+1} is independent of the σ -algebra \mathcal{J}_k . Then we have that:

(2.6)
$$E(t_{k+1}|\mathcal{J}_k) = E(I[U_{k+1} < \hat{\pi}_{k+1}]|\mathcal{J}_k) = \hat{\pi}_{k+1}$$

Let τ_k, ν_k be defined as the random variables which give the stage when the k^{th} observation from treatment A, respectively B, is taken. More explicitly, after the allocation is done we observe from treatment A the sequence of observations

 $Y^A_{\tau_1}, Y^A_{\tau_2}, \ldots$ and respectively from treatment B the sequence $Y^B_{\nu_1}, Y^B_{\nu_2}, \ldots$

Note that specifying the sequence $\tau_1, \nu_1, \tau_2, \nu_2, \ldots$ is equivalent to giving the sequence of the treatment indicators t_1, t_2, \ldots .

Indeed, τ_i, ν_i can be defined inductively as

$$au_0 = 0 \quad au_i = \inf\{k > au_{i-1} | t_k = 1\}, \quad \forall i \ge 1$$

$$\nu_0 = 0 \quad \nu_i = \inf\{k > \nu_{i-1} | t_k = 0\}, \quad \forall i \ge 1$$

and conversely,

$$egin{aligned} t_{ au_i} &= 1, \quad orall \, i \geq 1 \ t_{
u_i} &= 0, \quad orall \, i \geq 1 \ . \end{aligned}$$

The condition for the design that the treatment allocation for the k^{th} patient is based on the responses from the first k - 1 patients and the covariate of the k^{th} patient can be expressed as follows:

$$\{\tau_i = k\} \in \mathcal{G}_{k-1}, \{\nu_i = k\} \in \mathcal{G}_{k-1} \quad \forall i, k \ge 1$$

or equivalently,

$$(2.8) t_k \in \mathcal{G}_{k-1}$$

2.3 The Distribution of the Responses under Adaptive Design

We will show in this section that although the covariates can introduce a complicated dependence structure in the adaptive design, the marginal distribution of the responses will be unaffected. Theorem 2.1 from Melfi and Page(2000) says that in the case when the covariates are not considered in the model and as long as the pair (Y_k^A, Y_k^B) is independent of the σ -algebra \mathcal{F}_{k-1} then the allocated sequence inherit the distribution and independence structure of the original sequence.

In our case we do not get the independence, but we can show that the original distribution is inherited by the allocated sequence.

Theorem 2.3.1. Let $(Y_k^A, Y_k^B, X_k)_{k\geq 1}$ be a sequence of *i.i.d.* random vectors. Assume that the conditional distributions of Y_k^A, Y_k^B are given by equation (2.1) and that X_k is a discrete random vector.

Assume that there is a filtration $\{\mathcal{F}_k\}_{k\geq 1}$ such that $(Y_k^A, Y_k^B, X_k)_{k\geq 1}$ is \mathcal{F}_k measurable for all $k \geq 1$. Let $\{\mathcal{G}_k\}_{k\geq 1}$ be the filtration of the σ -algebra introduced by equation (2.4). Assume that (Y_k^A, Y_k^B) and \mathcal{G}_{k-1} are conditionally independent given X_k .

Let $\{\tau_k\}_{k\geq 1}$ and $\{\nu_k\}_{k\geq 1}$ be two sequences of positive, increasing, integer-valued, a.s. finite random variables which satisfy condition (2.7) and also satisfy

$$(2.9) P(\tau_i = \nu_j) = 0, \quad \forall i, j \ge 1$$

(2.10)
$$\forall k \geq 1, \exists i \text{ such that } \tau_i = k \text{ or } \nu_i = k.$$

Then the conditional distributions of the allocated sequence are given by the same

original distributions:

(2.11)
$$Y_{\tau_i}^A \mid X_{\tau_i} = x \sim F_x^A, \quad \forall i \ge 1$$
$$Y_{\tau_i}^B \mid X_{\tau_i} = x \sim F_x^B, \quad \forall i \ge 1.$$

Before we proceed to the proof of the theorem we remark that the condition (2.9) is the expression of the initial assumption that a patient can receive only one treatment and the condition (2.10) excludes the trivial case when some patients do not receive any treatment at all.

Proof. We'll prove the result only for $Y_{\tau_i}^A$, the proof for $Y_{\nu_i}^B$ is similar.

We want to prove that for any measurable set D,

$$P(Y_{\tau_i}^A \in D | X_{\tau_i} = x) = F_x^A(D) .$$

We have that

$$P(Y_{\tau_i}^A \in D | X_{\tau_i} = x) = \sum_{m \ge 1} P(Y_{\tau_i}^A \in D, \tau_i = m | X_{\tau_i} = x) .$$

Recall from condition (2.7) that $\{\tau_i = m\} \in \mathcal{G}_{m-1}$ for all $i, m \ge 1$ and using the hypothesis that Y_m^A and \mathcal{G}_{m-1} are conditionally independent given X_m we obtain that:

$$P(Y_m^A \in D, \tau_i = m \mid X_m = x) = P(Y_m^A \in D \mid X_m = x)P(\tau_i = m \mid X_m = x)$$

Then in the case when $P(X_{\tau_i} = x) > 0$

$$\begin{split} P(Y_{\tau_{i}}^{A} \in D, \tau_{i} = m \mid X_{\tau_{i}} = x) &= \frac{P(Y_{\tau_{i}}^{A} \in D, \tau_{i} = m, X_{\tau_{i}} = x)}{P(X_{\tau_{i}} = x)} \\ &= \frac{P(Y_{m}^{A} \in D, \tau_{i} = m, X_{m} = x)}{P(X_{\tau_{i}} = x)} \\ &= \frac{P(Y_{m}^{A} \in D, \tau_{i} = m \mid X_{m} = x)P(X_{m} = x)}{P(X_{\tau_{i}} = x)} \\ &= \frac{P(Y_{m}^{A} \in D \mid X_{m} = x)P(\tau_{i} = m \mid X_{m} = x)P(X_{m} = x)}{P(X_{\tau_{i}} = x)} \\ &= \frac{F_{x}^{A}(D)P(\tau_{i} = m \mid X_{m} = x)P(X_{m} = x)}{P(X_{\tau_{i}} = x)} \\ &= \frac{F_{x}^{A}(D)P(\tau_{i} = m, X_{m} = x)}{P(X_{\tau_{i}} = x)} \\ &= \frac{F_{x}^{A}(D)P(\tau_{i} = m, X_{\tau_{i}} = x)}{P(X_{\tau_{i}} = x)} \\ &= F_{x}^{A}(D)P(\tau_{i} = m \mid X_{\tau_{i}} = x) \\ &= F_{x}^{A}(D)P(\tau_{i} = m \mid X_{\tau_{i}} = x) . \end{split}$$

Combining the relations together, we get indeed the equality required.

2.3.1 A counterexample

We will show in this section why the independence structure of covariates is not generally preserved in the case of an adaptive design. For this we will look at a counterexample, an adaptive design satisfying the properties required in the beginning of the chapter but for which the sequence of the observed covariates is not independent.

More precisely, keeping the same notation as before, we will show that the sequence $X_{\tau_1}, X_{\tau_2}, \ldots$ is not independent and similarly for $X_{\nu_1}, X_{\nu_2}, \ldots$

Assume the following:

• $X_1, X_2, ...$ are i.i.d. Bernoulli with $P(X_1 = 1) = P(X_1 = 0) = 0.5;$

• For the responses assume the following distributions:

$$P(Y_k^A = 1 | X_k = 1) = 1 - P(Y_k^A = 0 | X_k = 1) = .75$$

$$P(Y_k^A = 1 | X_k = 0) = 1 - P(Y_k^A = 0 | X_k = 0) = .25$$

$$P(Y_k^B = 1 | X_k = 1) = 1 - P(Y_k^B = 0 | X_k = 1) = .25$$

$$P(Y_k^B = 1 | X_k = 1) = 1 - P(Y_k^B = 0 | X_k = 1) = .75$$

We associate the value 1 for the response with a successful treatment.

• Assume the following iterative design:

$$P(t_1 = 1) = P(t_1 = 0) = 0.5$$

(in the first stage we assume the equipoise state so there is no preferred treatment and the covariate is ignored).

At each stage $k \geq 2$ let $N_k^A = \sum_{i=1}^k t_k$ be the number of allocations to treatment A up to stage k. Let $k^- = \tau_{N_{k-1}^A} = \max_{1 \leq i < k} \{i | t_i = 1\}$ be the last time before the k^{th} patient when we allocated treatment A.

- If $k^- = 0$ which corresponds to the case that there were no allocations to treatment A before the k^{th} patient then just allocate with equal probability disregarding again the covariate:

$$P(t_k = 1) = P(t_k = 0) = 0.5$$

– If $k^- > 0$ then allocate according to the following rule:

$$P(t_{k} = 1|X_{k} = 1) = 1 - P(t_{k} = 0|X_{k} = 1) = \begin{cases} 0.9 & \text{if } X_{k^{-}} = 1, Y_{k^{-}}^{A} = 1\\ 0.3 & \text{if } X_{k^{-}} = 1, Y_{k^{-}}^{A} = 0\\ 0.7 & \text{if } X_{k^{-}} = 0, Y_{k^{-}}^{A} = 1\\ 0.4 & \text{if } X_{k^{-}} = 0, Y_{k^{-}}^{A} = 0 \end{cases}$$
$$P(t_{k} = 1|X_{k} = 0) = 1 - P(t_{k} = 1|X_{k} = 0) = \begin{cases} 0.7 & \text{if } X_{k^{-}} = 1, Y_{k^{-}}^{A} = 1\\ 0.4 & \text{if } X_{k^{-}} = 1, Y_{k^{-}}^{A} = 1\\ 0.4 & \text{if } X_{k^{-}} = 1, Y_{k^{-}}^{A} = 0\\ 0.9 & \text{if } X_{k^{-}} = 0, Y_{k^{-}}^{A} = 1\\ 0.3 & \text{if } X_{k^{-}} = 0, Y_{k^{-}}^{A} = 0 \end{cases}$$

We show now that $X_{\tau_1}, X_{\tau_2}, \ldots$ are not independent.

$$P(X_{\tau_1} = 1) = \sum_{k \ge 1} P(X_{\tau_1} = 1, \tau_1 = k)$$

= $\sum_{k \ge 1} P(X_k = 1, t_1 = 0, \dots, t_{k-1} = 0, t_k = 1)$
= $\sum_{k \ge 1} P(t_k = 1 | X_k = 1, t_1 = 0, \dots, t_{k-1} = 0)$
 $\cdot P(X_k = 1) \cdot P(t_1 = 0, \dots, t_{k-1} = 0)$
= $\sum_{k \ge 1} \frac{1}{2} \cdot \frac{1}{2} \cdot P(t_1 = 0, \dots, t_{k-1} = 0)$;

and similarly,

$$P(X_{\tau_1} = 0) = \sum_{k \ge 1} P(X_{\tau_1} = 0, \tau_1 = k)$$

= $\sum_{k \ge 1} P(X_k = 0, t_1 = 0, \dots, t_{k-1} = 0, t_k = 1)$
= $\sum_{k \ge 1} P(t_k = 1 | X_k = 0, t_1 = 0, \dots, t_{k-1} = 0)$
 $\cdot P(X_k = 0) \cdot P(t_1 = 0, \dots, t_{k-1} = 0)$
= $\sum_{k \ge 1} \frac{1}{2} \cdot \frac{1}{2} \cdot P(t_1 = 0, \dots, t_{k-1} = 0)$.

And since $P(X_{\tau_1} = 1) + P(X_{\tau_1} = 0) = 1$ we obtain that

$$P(X_{\tau_1} = 1) = P(X_{\tau_1} = 0) = 0.5$$

On the other hand,

$$P(X_{\tau_2} = 1) = \sum_{k \ge 1} \sum_{k > j \ge 1} P(X_{\tau_2} = 1, \tau_2 = k, \tau_1 = j)$$

= $\sum_{k \ge 1} \sum_{k > j \ge 1} P(X_k = 1, t_k = 1, t_{k-1} = 0, \dots, t_{j+1} = 0, t_j = 1, t_{j-1} = 0, \dots, t_1 = 0)$
 $\triangleq \sum_{k \ge 1} \sum_{k > j \ge 1} P(A_k) ,$

where A_k denotes the event:

$$A_{k} = \{X_{k} = 1, t_{k} = 1, t_{k-1} = 0, \dots, t_{j+1} = 0, t_{j} = 1, t_{j-1} = 0, \dots, t_{1} = 0\}.$$

Let $B_{11,j} = \{Y_j = 1, X_j = 1\}, B_{10,j} = \{Y_j = 1, X_j = 0\}, B_{01,j} = \{Y_j = 0, X_j = 1\}$

and $B_{00,j} = \{Y_j = 0, X_j = 0\}$. Then:

$$P(X_{\tau_2} = 1) = \sum_{k \ge 1} \sum_{k > j \ge 1} \sum_{\substack{\alpha = 0, 1 \\ \beta = 0, 1}} P(A_k \cap B_{\alpha \beta, j})$$
$$= I_{11} + I_{10} + I_{01} + I_{00} .$$

We'll show the computations just for the first summand:

$$\begin{split} I_{11} &= P(X_k = 1, t_k = 1, t_{k-1} = 0, \dots, t_{j+1} = 0, t_j = 1, B_{11,j}, t_{j-1} = 0, \dots, t_1 = 0) \\ &= P(t_k = 1 | X_k = 1, t_{k-1} = 0, \dots, t_{j+1} = 0, t_j = 1, B_{11,j}, t_{j-1} = 0, \dots, t_1 = 0) \\ &\cdot P(X_k = 1) \\ &\cdot P(t_{k-1} = 0, \dots, t_{j+1} = 0 | t_j = 1, B_{11,j}, t_{j-1} = 0, \dots, t_1 = 0) \\ &\cdot P(Y_j = 1 | t_j = 1, X_j = 1, t_{j-1} = 0, \dots, t_1 = 0) \\ &\cdot P(t_j = 1 | X_j = 1, t_{j-1} = 0, \dots, t_1 = 0) \\ &\cdot P(X_j = 1) \\ &\cdot P(t_{j-1} = 0, \dots, t_1 = 0) \\ &= 0.9 \cdot 0.5 \cdot (0.1)^{k-j-1} \cdot 0.75 \cdot 0.5 \cdot 0.5 \cdot 0.5^{j-1} . \end{split}$$

In a similar way we obtain that:

$$I_{10} = 0.7 \cdot 0.5 \cdot (0.3)^{k-j-1} \cdot 0.25 \cdot 0.5 \cdot 0.5 \cdot 0.5^{j-1}$$
$$I_{01} = 0.3 \cdot 0.5 \cdot (0.7)^{k-j-1} \cdot 0.25 \cdot 0.5 \cdot 0.5 \cdot 0.5^{j-1}$$
$$I_{00} = 0.4 \cdot 0.5 \cdot (0.6)^{k-j-1} \cdot 0.75 \cdot 0.5 \cdot 0.5 \cdot 0.5^{j-1}$$
Also we can to observe that:

$$P(X_{\tau_2} = 1, X_{\tau_1} = 1) = \sum_{k \ge 1} \sum_{k > j \ge 1} P(X_{\tau_2} = 1, \tau_2 = k, X_{\tau_1} = 1, \tau_1 = j)$$

$$= \sum_{k \ge 1} \sum_{k > j \ge 1} P(X_k = 1, t_k = 1, t_{k-1} = 0, \dots, t_{j+1} = 0, X_j = 1, t_j = 1, t_{j-1} = 0, \dots, t_1 = 0)$$

$$= \sum_{k \ge 1} \sum_{k > j \ge 1} P(A_k \cap \{X_j = 1\})$$

$$= \sum_{k \ge 1} \sum_{k > j \ge 1} P(A_k \cap \{Y_j = 1\} \cap \{X_j = 1\} + P(A_k \cap \{Y_j = 0\} \cap \{X_j = 1\}))$$

$$= I_{11} + I_{01}.$$

Since from the above relations is clear that $I_{11} + I_{01} \neq I_{00} + I_{10}$ then it follows that:

$$P(X_{\tau_1} = 1)P(X_{\tau_2} = 1) = \frac{1}{2}(I_{11} + I_{01} + I_{10} + I_{00}) \neq \frac{1}{2}(I_{11} + I_{01}) = P(X_{\tau_2} = 1, X_{\tau_1} = 1).$$

which proves the property that the sequence $(X_{\tau_k})_{k\geq 1}$ is not independent.

Chapter 3

Asymptotic Results for Adaptive Designs with Covariates

3.1 Introduction

In this section we will introduce a specific parametric model for the problem we formulated in the beginning of the previous chapter. Thus we will concentrate now on the study of the behavior of the parameters rather than the study of distribution of responses themselves. We need a different approach from Melfi and Page (2000) to derive the asymptotic results because in their case, the adaptive design produced an i.i.d. sequence for the responses, result which is not available in the case of the adaptive design with covariates.

We will work under the same assumptions mentioned in the second chapter, namely:

- Patients arrive sequentially.
- Each patient can be assigned to either one of the two treatments and will be assigned to exactly one of them.

- The response from patients is observed immediately, prior to the assignment of the treatment to the next patient.
- The response is possible dependent on a vector of covariates. The covariates are observed prior to the treatment assignment.

Our goal is to show how a response-adaptive allocation where the treatment allocation depends on the previous responses and the current covariate of the patient influences the asymptotic results for estimators of the parameters of the model we propose.

As pointed out in the first chapter much of the work done in the area of the adaptive designs with covariates concentrated on the approach of forming strata by considering all possible combinations of levels of relevant covariates and then independently use different adaptive schemes within each stratum to allocate patients. But the approach that we will take in this chapter comes naturally whenever we can model a response which is a function of covariates. Specifically, we will consider a generalized linear model for linking the response with the covariates and the treatment, and then we will concentrate on studying the behavior of the estimators of the parameter vector of coefficients under a general adaptive allocation which satisfies the above mentioned assumptions.

We will prove consistency and asymptotic normality results for the estimators of the parameter vector of coefficients under some conditions imposed on the model.

3.2 A Generalized Linear Model Relating the Response, the Adaptive Design and the Covariates

In what follows we introduce the model and notation that we'll need through the rest of the chapter. We will follow as much as possible the notation introduced in the second chapter and we will extend the definitions previously used when needed.

As they were introduced in the previous chapter, for each $k \ge 1$ let t_k be the treatment allocation to the k^{th} patient:

$$t_{k} = \begin{cases} 1 & \text{if } k^{th} \text{ patient is assigned to treatment A} \\ 0 & \text{if } k^{th} \text{ patient is assigned to treatment B} \end{cases}$$

Also recall that $X_k = (X_{k1}, \ldots, X_{kp})$ denotes, as before, the *p*-dimensional covariate vector of the k^{th} patient.

Let Y_k be the observed response of the k^{th} patient. Using the notation Y_k^A and Y_k^B introduced in the previous chapter we can define Y_k as:

(3.1)
$$Y_k = Y_k^A I[t_k = 1] + Y_k^B I[t_k = 0]$$

where we denote by $I(\cdot)$ the usual indicator function.

1

As will be seen below, we don't have to make any longer the difference in the notation between response from treatment A and respectively treatment B, Y_k^A and Y_k^B . This is because with the new approach, we will consider the response Y_k in the context of a generalized linear model, with the independent variables being the covariate X_k and treatment t_k and with possible interactions.

Assume, as before, the existence of an increasing sequence of σ -algebra $\{\mathcal{F}_k\}_{k\geq 1}$

defined as

(3.2)
$$\mathcal{F}_k := \sigma(Y_i^A, Y_i^B, X_i, t_i, U_i : 1 \le i \le k) .$$

As we discussed in chapter 2, $(U_k)_{k\geq 1}$ denotes a sequence of i.i.d. random variables uniform distributed on (0, 1) and independent of the other variables in the model (Y_k^A, Y_k^B, X_k) . They may be used for randomizing the allocation treatment as in the construction given by (2.2).

Consider also the increasing sequence of σ -algebras $\{\mathcal{G}_k\}_{k\geq 1}$ defined as in equation (2.4):

$$\mathcal{G}_k := \mathcal{F}_k \vee \sigma(X_{k+1}) \vee \sigma(U_{k+1})$$
.

Recall that our goal is to make the allocation decision for the k^{th} patient based on the responses from the first k - 1 patients and the covariate of the k^{th} patient. Hence in this case the condition can be expressed as

$$(3.3) t_k \in \mathcal{G}_{k-1}$$

Before we introduce our model, recall an overview of the generalized linear model. The generalized linear model is characterized by a random part and a structural part. The random part consists of the independent observations $(Y_i)_{i\geq 1}$ whose distributions are members of the same exponential family, given by the generic density function:

$$f(y_i| heta_i,\phi,w_i) = \exp\left\{rac{y_i\cdot heta_i-b(heta_i)}{\phi}\cdot w_i + c(y_i,\phi,w_i)
ight\}$$

where

- θ_i is the so-called natural parameter
- ϕ is the dispersion parameter

- w_i is a known weight
- $c(\cdot, \cdot, \cdot)$ is a function of y_i , dispersion ϕ and weight w_i .

The structural part of the model specifies the linear relationship between a vector of covariates $x_i = (x_{i1}, \ldots, x_{iq})$ and the expectation of the response, $E(Y_i) = \mu_i$:

$$g(\mu_i) = \eta_i = x_i \beta = \sum_{j=1}^q x_{ij} \cdot \beta_j$$
.

The invertible function $g(\cdot)$, common to all observations, is called the link function and β is the parameter vector of the model. Each exponential distribution is associated with a canonical link function, although this is not necessarily the best link for modeling.

The more familiar linear regression model and logistic regression model can be both viewed as particular examples of the generalized linear model. For the case when the random part has binomial distribution and the canonical link is the logit function, $g(y) = \text{logit}(y) = \log(\frac{y}{1-y})$ we obtain from the generalized linear model the logistic regression model. In the case when the random part has normal distribution and the canonical link is the identity function g(y) = y we obtain the linear regression model.

In our case we construct the generalized linear model as follows. The structural part is given by the following formula:

$$(3.4) E(Y_k|\mathcal{F}_{k-1} \lor \sigma(X_k, t_k)) = g(X_k\beta_1 + t_k\beta_2 + X_k\beta_3 t_k)$$

where

• g is the link function associated with the distribution of the response Y_k

• β_1, β_3 are p -dimensional column vectors and β_2 is scalar:

$$\beta'_1 = (\beta_{11}, \beta_{12}, \dots, \beta_{1p})$$

$$\beta'_3 = (\beta_{31}, \beta_{32}, \dots, \beta_{3p}) .$$

We will use from now on the notation M' to denote the transpose of a matrix M.

We used in relation (3.4) the σ -algebra $\mathcal{F}_{k-1} \vee \sigma(X_k, t_k)$ to emphasize the relationship between the response Y_k and X_k, t_k . But since $t_k \in \mathcal{G}_{k-1}$ by assumption (3.3) and using the definition (2.4) of the σ -algebra \mathcal{G}_{k-1} we may use as well the following definition for the structural part of our model:

$$(3.5) E(Y_k|\mathcal{G}_{k-1}) = g(X_k\beta_1 + t_k\beta_2 + X_k\beta_3 t_k)$$

Consider the errors ϵ_k for the model:

(3.6)
$$\epsilon_{k} = Y_{k} - E(Y_{k}|\mathcal{G}_{k-1}) .$$

We will assume that the errors ϵ_k are independent.

In a more explicit form, the systematic part of the model can be rewritten as:

(3.7)
$$E(Y_k|\mathcal{G}_{k-1}) = g\left(\sum_{j=1}^p X_{kj}\beta_{1j} + t_k\beta_2 + \sum_{j=1}^p X_{kj}\beta_{3j}t_k\right)$$

The coefficient parameter of the model is given by the (2p + 1) -dimensional column vector

$$eta'=(eta_1',eta_2,eta_3')$$
 .

In some cases it will be useful to refer to the regressor vector of the model (3.4)

as a whole. We introduce the following notation:

$$(3.8) X_k^* = \begin{pmatrix} X_k, t_k, X_k \cdot t_k \end{pmatrix} .$$

The innovation of our model is that we introduce in the structural part of generalized linear model the dependence on the treatment where the treatment allocation follows an adaptive design as defined by relation (3.3).

By considering the generalized linear model we allow either continuous or categorical responses Y_k . For the continuous responses, normal errors and the case when the link function is the identity function g(y) = y, we get back to the usual linear model.

For now we don't assume yet any special conditions for the errors, except independence, but as we will see later in the chapter we will need to enforce some regularity conditions on them to prove the consistency and the asymptotic properties for the estimators of the parameters of the model.

Our next goal is to define some estimators for the coefficient parameters of the model (3.4) and then to study their behavior. The most natural estimators that arise are the maximum quasi-likelihood estimators.

The maximum quasi-likelihood estimators are defined in connection with the socalled quasi-likelihood functions introduced by Wedderburn (1974) and McCullagh & Nelder (1989). In general, if we have independent observations $(y_i)_{i=1,...,n}$ with expectations μ_i and variances $V(\mu_i)$ then the quasi-likelihood function $K(y_i, \mu_i)$ is defined by the relation:

$$rac{\partial K(y_i,\mu_i)}{\partial \mu_i} = rac{y_i - \mu_i}{V(\mu_i)} \; .$$

Assume that for each observation, μ_i is a function of parameters β_1, \ldots, β_q . It

can be proved that

$$E\left(rac{\partial K}{\partialeta_i}
ight)=0 \quad ext{for all } i=1,\ldots,q.$$

Then the maximum quasi-likelihood estimators are the solutions from equating $\partial K/\partial \beta_i$ with its expectation, zero:

$$\sum_{j=1}^{n} \frac{\partial K(y_j, \mu_j(\beta_1, \dots, \beta_q))}{\partial \beta_i} = 0 \quad \text{for all } i = 1, \dots, q.$$

Let $\hat{\beta}'_n = (\hat{\beta}'_{1,n}, \hat{\beta}'_{2,n}, \hat{\beta}'_{3,n})$ be the maximum quasi-likelihood estimator of the model (3.4). Note again that the estimators $\hat{\beta}_{1,n}, \hat{\beta}_{3,n}$ are *p*-dimensional column vectors and $\hat{\beta}_{1,n}$ is scalar. Explicitly in our case $\hat{\beta}_n$ is the solution of the following equations:

(3.9)
$$\sum_{i=1}^{n} x_i [y_i - g(x_i \hat{\beta}_{1,n} + \hat{\beta}_{2,n} \cdot t_i + x_i \hat{\beta}_{3,n} \cdot t_i)] = 0$$
$$\sum_{i=1}^{n} t_i [y_i - g(x_i \hat{\beta}_{1,n} + \hat{\beta}_{2,n} \cdot t_i + x_i \hat{\beta}_{3,n} \cdot t_i)] = 0$$
$$\sum_{i=1}^{n} x_i t_i [y_i - g(x_i \hat{\beta}_{1,n} + \hat{\beta}_{2,n} \cdot t_i + x_i \hat{\beta}_{3,n} \cdot t_i)] = 0.$$

We may remark that in the case of linear models, when the link function $g(\cdot)$ is the identity function, equations (3.9) become the normal equations and the maximum quasi-likelihood estimator becomes the usual least squares estimator.

We will prove in the next sections that under some assumptions we can derive consistency and asymptotic normality properties for the maximum quasi-likelihood estimators defined as solutions to equations (3.9). In return the results allow us to analyze predictors for the response Y_k from treatments A and B and eventually to compare them. Let \mathbf{Z}_n be the model (3.4) design matrix, that is

$$(3.10) \mathbf{Z}_{n} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1p} & X_{11} \cdot t_{1} & X_{12} \cdot t_{2} & \dots & X_{1p} \cdot t_{1} & t_{1} \\ X_{21} & X_{22} & \dots & X_{2p} & X_{21} \cdot t_{2} & X_{22} \cdot t_{2} & \dots & X_{2p} \cdot t_{2} & t_{2} \\ \vdots & & & & & \\ X_{n1} & X_{n2} & \dots & X_{np} & X_{n1} \cdot t_{n} & X_{n2} \cdot t_{n} & \dots & X_{np} \cdot t_{n} & t_{n} \end{pmatrix}$$

We will pay special attention to the eigenvalues of the matrix $\mathbf{Z}'_{n}\mathbf{Z}_{n}$.

Let $\lambda_{min}(n)$ and $\lambda_{max}(n)$ be the minimum and respectively the maximum eigenvalues of $\mathbf{Z}'_{n}\mathbf{Z}_{n}$. The assumptions for the results we will prove next depend on the regularity of the design matrix \mathbf{Z}_{n} and this will be expressed in terms of the above defined eigenvalues.

3.3 A Consistency Result

We want to show in this section that the maximum quasi-likelihood estimator, $\hat{\beta}_n$, is strongly consistent for the model parameter, β . In order to prove the result we need some limit constraints for the errors of the model, ϵ_n , and also we will rule out the asymptotically ill-conditioned design matrices, \mathbf{Z}_n .

Theorem 2 from Chen, Hu and Ying (1999) and Theorem 1 from Lai and Wei (1982) address the consistency problem for an adaptive design for a generalized linear model and respectively a linear regression model. But in their case the design does not allow for the treatment allocation t_k to be dependent on the covariate X_k . But in our case our main design goal is to make the treatment allocation, t_k dependent not only on the past information \mathcal{F}_{k-1} but also on the covariate value of the current patient, X_k . We will mirror the results mentioned before but using the new σ -algebras required by our design.

In what follows we will denote by the norm $\|\cdot\|_2$ the usual Euclidean norm, that

is for an *m*-dimensional vector $x = (x_1, x_2, \ldots, x_m)$,

$$||x||_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_m^2}$$
.

Theorem 3.3.1. Assume that we have a generalized linear model as specified by the model (3.4) (3.6) with an adaptive design satisfying relations (3.2) and (3.3). Assume the following conditions are satisfied:

(3.11) g is continuously differentiable with positive derivative function

(3.12)
$$\sup_{i\geq 1} \|X_i\|_2 < \infty \, a.s. \; ,$$

(3.13)
$$\lim_{n \to \infty} \lambda_{min}(n) / \log \lambda_{max}(n) = \infty \ a.s. ,$$

(3.14)
$$\sup_{i\geq 1} E(|\epsilon_i|^{\alpha}|\mathcal{G}_{i-1}) < \infty \ a.s. \ for \ some \ \alpha > 2 \ .$$

Then the estimator $\hat{\beta}_n$ is strongly consistent, and

(3.15)
$$\|\hat{\beta}_n - \beta\|_2 = O((\log \lambda_{max}(n)/\lambda_{min}(n))^{1/2}) a.s.$$

Let recall first the theorem proved by Chen, Hu and Ying (1999) in their paper.

Theorem 3.3.2 (Chen, Hu and Ying (1999)). Consider a generalized linear model for the pairs $(y_k, x_k)_{k\geq 1}$ given by:

$$(3.16) E(y_k|\mathcal{F}_{k-1}) = g(\beta \cdot x_k)$$

where $(\mathcal{F}_k)_{k\geq 1}$ is a σ -filtration such that $y_k \in \mathcal{F}_k$ and $x_k \in \mathcal{F}_{k-1}$. Let the errors ϵ_k be defined as:

$$\epsilon_k = y_k - E(y_k | \mathcal{F}_{k-1})$$

Let λ_{\min} , λ_{\max} be the minimum and the maximum eigenvalues of the information matrix $\sum_{i=1}^{n} x_i x'_i$. Let $\hat{\beta}_n$ be the maximum quasi-likelihood estimator of β from model (3.16). Assume the following conditions are satisfied:

$$(3.17) g in s continuously differentiable with positive derivative function$$

(3.18)
$$\sup_{i\geq 1} \|x_i\|_2 < \infty \, a.s.$$

(3.19)
$$\lim_{n \to \infty} \lambda_{min}(n) / \log \lambda_{max}(n) = \infty a.s.$$

(3.20)
$$\sup_{i\geq 1} E(|\epsilon_i|^{\alpha}|\mathcal{F}_{i-1}) < \infty \ a.s. \ for \ some \ \alpha > 2$$

Then

(3.21)
$$\|\hat{\beta}_n - \beta\|_2 = O((\log \lambda_{max}(n) / \lambda_{min}(n))^{1/2}) a.s.$$

As we pointed out before, the implementation of the new σ -algebras \mathcal{G}_k will solve the problem for our design.

Proof of theorem 3.3.1. From the construction of the σ -algebras \mathcal{G}_k we have that t_k is \mathcal{G}_{k-1} -measurable and from the construction itself of the filtration, X_k is \mathcal{G}_{k-1} -measurable and Y_k is \mathcal{G}_k -measurable. The model (3.4) becomes:

$$E(Y_k|\mathcal{G}_{k-1}) = g(\beta_1 \cdot X_k + \beta_2 \cdot t_k + \beta_3 \cdot X_k \cdot t_k)$$

We can apply now theorem 3.3.2 with \mathcal{F}_k replaced by \mathcal{G}_k and covariate matrix $\sum_{i=1}^n x_i x'_i$ replaced by $\mathbf{Z}'_n \mathbf{Z}_n$.

We observe that conditions (3.17) (3.19) (3.20) are satisfied by the equivalent conditions (3.11) (3.13) (3.14) from theorem 3.3.1. To see that also condition (3.18) is satisfied we can use the inequality

$$\|X_{k}^{*}\|_{2}^{2} = \left\| \begin{pmatrix} X_{k}, t_{k}, X_{k} \cdot t_{k} \end{pmatrix} \right\|_{2}^{2}$$
$$\leq 2 \cdot \|X_{k}\|_{2}^{2} + 1$$

because $t_k \in \{0, 1\}$ for all $k \ge 1$. Now using (3.12) it follows that indeed condition (3.18) is satisfied. Hence the conclusion (3.21) of Theorem 3.3.2 which is just exactly what we need for (3.15), too.

We will look now at the particular case when the covariate is 1-dimensional, that is, p = 1. Then we can find explicitly the eigenvalues involved in the condition (3.13) from Theorem 3.3.1 and we will be able to prove the desired limit under some verifiable conditions.

Proposition 3.3.3. Let covariate $(x_i)_{i\geq 1}$ and treatment $(t_i)_{i\geq 1}$ sequences be such that:

$$\begin{split} &\sum_{i\geq 1} x_i^2 \to \infty \, a.s. \\ &\sum_{i\geq 1} t_i^2 \to \infty \, a.s. \\ &\frac{\left(\sum_{i\geq 1} x_i t_i\right)^2}{\sum_{i\geq 1} x_i^2 \cdot \sum_{i\geq 1} t_i^2} < 1 \, a.s. \end{split}$$

Let

$$\mathbf{Z}'_{n} = \begin{pmatrix} x_{1} & x_{2} & \dots & x_{n} \\ \\ t_{1} & t_{2} & \dots & t_{n} \end{pmatrix}$$

Then

$$\lim_{n \to \infty} \frac{\lambda_{min}(n)}{\log \lambda_{max}(n)} \to \infty \, a.s.$$

where $\lambda_{\min}(n)$, $\lambda_{\max}(n)$ are the minimum, respective maximum eigenvalue of the matrix $\mathbf{Z}'_{n}\mathbf{Z}_{n}$.

Proof. Let $m_1 = \sum_{i=1}^n x_i^2$; $m_2 = \sum_{i=1}^n x_i t_i$ and $m_3 = \sum_{i=1}^n t_i^2$. We omitted the index n to make the notation easier. It can be readily proved that

$$\lambda_{min} = \frac{1}{2} \left(m_1 + m_3 - \sqrt{(m_1 - m_3)^2 + 4m_2^2} \right)$$
$$\lambda_{max} = \frac{1}{2} \left(m_1 + m_3 + \sqrt{(m_1 - m_3)^2 + 4m_2^2} \right)$$

Because of Cauchy inequality we have that $m_2^2 \leq m_1 m_3$. Hence $\lambda_{min} \geq 0$. Moreover,

$$\lambda_{min} = 2 \cdot \frac{1 - \frac{m_2^2}{m_1 m_3}}{\frac{1}{m_1} + \frac{1}{m_3} + \sqrt{\left(\frac{1}{m_1} - \frac{1}{m_3}\right)^2 + \frac{m_2^2}{m_1 m_3}\frac{4}{m_1 m_3}}}$$

Let

$$l:=\frac{m_2^2}{m_1m_3}$$

and from hypothesis, at the limit, l < 1 a.s.. Using also the hypothesis that $m_1, m_3 \rightarrow \infty a.s.$ we have that as $n \rightarrow \infty$,

$$\lambda_{min} \to 2 \cdot rac{1-l}{0+0+\sqrt{(0-0)^2+l\cdot 0}} a.s.$$

therefore as $n \to \infty$, $\lambda_{min} \to \infty a.s.$

Let q be a constant such that

$$q > \frac{1 + \sqrt{l}}{1 - \sqrt{l}}$$

_

Then using some algebra we can prove that

$$q \cdot \lambda_{min} > \lambda_{max}$$

Then we have that

$$\frac{\lambda_{\min}}{\log \lambda_{\max}} = \frac{\lambda_{\min}}{\log \lambda_{\min}} \cdot \frac{\log \lambda_{\min}}{\log \lambda_{\max}}$$

and

$$\frac{\log \lambda_{min}}{\log(q \cdot \lambda_{min})} < \frac{\log \lambda_{min}}{\log \lambda_{max}} < 1$$

Since $\lambda_{min} \to \infty a.s.$ we obtain that $\log \lambda_{min} / \log \lambda_{max} \to 1 a.s.$ and therefore $\lambda_{min} / \log \lambda_{max} \to \infty a.s.$ when $n \to \infty$.

3.4 An Asymptotic Normality Result

We showed in the previous section that the maximum quasi-likelihood estimator, $\hat{\beta}_n$, is consistent under assumptions (3.11) - (3.14). Therefore, under the adaptive design and the model considered, we know that at least asymptotically we get the correct estimators for the coefficient parameter β . Our next goal is to study the asymptotic normality behavior of the maximum likelihood estimators. We will prove the asymptotic normality result in the case of the linear regression when also the maximum quasi-likelihood estimator becomes the usual least square estimator. Unfortunately in the literature there is not any similar result for the generalized linear model. Because we need the variance for the asymptotic normal distribution we will need an extra asymptotic limit condition on the errors of model (3.4).

Recall that Theorem 3 from Lai and Wei (1982) addresses also the asymptotic normality problem but the design in their case does not permit the choice of the treatment to be dependent on the covariate of the current patient. We will extend the result they obtained for their particular case to the design that we presented in the introduction of the chapter.

Theorem 3.4.1. Suppose that for the model (3.4) with the identity link function, $g(\cdot) = \cdot$, the errors ϵ_k satisfy condition (3.14) and the following limit condition:

(3.22)
$$\lim_{i\to\infty} E(\epsilon_i^2 | \mathcal{G}_{i-1}) = \sigma^2 a.s.$$

Moreover there exists a non-random positive definite symmetric matrix \mathbf{B}_n for which

(3.23)
$$\lim_{n \to \infty} \mathbf{B}_n^{-1} (\mathbf{Z}_n' \mathbf{Z}_n)^{\frac{1}{2}} \xrightarrow{P} \mathbf{I}$$

(3.24)
$$\lim_{n \to \infty} \max_{1 \le i \le n} \|\mathbf{B}_n^{-1} X_i^*\|_2 \xrightarrow{P} 0$$

Then the estimator $\hat{\beta}_n$ has an asymptotically normal distribution:

(3.25)
$$(\mathbf{Z}'_{n}\mathbf{Z}_{n})^{\frac{1}{2}}(\hat{\beta}_{n}-\beta) \xrightarrow{D} \mathbf{N}(\mathbf{0},\sigma^{2}\mathbf{I})$$

Let recall first the theorem proved by Lai and Wei (1982) in their paper.

Theorem 3.4.2 (Lai and Wei (1982)). Suppose that for the regression model

$$(3.26) y_n = \beta_1 x_{n1} + \dots + \beta_p x_{np} + \epsilon_n ,$$

 ε_n is a martingale sequence with respect to an increasing sequence of σ -fields \mathcal{F}_n such that

(3.27)
$$\sup_{i\geq 1} E(|\epsilon_i|^{\alpha}|\mathcal{F}_{i-1}) < \infty \ a.s. \ for \ some \ \alpha > 2$$

(3.28)
$$\lim_{n\to\infty} E(\epsilon_i^2 | \mathcal{F}_{i-1}) = \sigma^2 a.s. \text{ for some constant } \sigma$$

Moreover, assume for each n that the design vector

(3.29)
$$\mathbf{x}_i = (x_{i1}, \dots, x_{ip})' \text{ is } \mathcal{F}_{i-1} - measurable$$

and that there exists a non-random positive definite \mathbf{B}_n for which

(3.30)
$$\lim_{n \to \infty} \mathbf{B}_n^{-1} (\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i)^{\frac{1}{2}} \xrightarrow{P} \mathbf{I}$$

(3.31)
$$\lim_{n \to \infty} \max_{1 \le i \le n} \|\mathbf{B}_n^{-1} \mathbf{x}_i\|_2 \xrightarrow{P} 0$$

Then the least square estimate $\hat{\beta}_n$ of β has an asymptotically normal distribution:

(3.32)
$$(\sum_{i=1}^{n} \mathbf{x}_{i} \mathbf{x}_{i}')^{\frac{1}{2}} (\hat{\beta}_{n} - \beta) \xrightarrow{D} \mathbf{N}(\mathbf{0}, \sigma^{2} \mathbf{I})$$

Proof of theorem 3.4.1. Again the crucial step is the definition of the σ -algebra \mathcal{G}_k we introduced by equation (2.4).

From that definition and from condition (3.3) we have that $X_k, t_k \in \mathcal{G}_{k-1}$, hence X_k^* is \mathcal{G}_{k-1} -measurable for all k so condition (3.29) is satisfied. From the equation (3.6) we have that the errors form already a martingale difference sequence. By conditions (3.22) and (3.14) we have satisfied conditions (3.27) and (3.28). Finally replacing the design vector \mathbf{x}_n by X_n^* and the covariate matrix $\sum_{i=1}^n x_i x_i'$ by $\mathbf{Z}'_n \mathbf{Z}_n$ we get that the conditions (3.30) and (3.31) are satisfied by conditions (3.23) and (3.24). The conclusion (3.25) follows from (3.32).

Chapter 4

Examples of Models with an Adaptive Allocation

4.1 Introduction

In the previous two chapters we proved some general theorems regarding, first, the distribution of the response from a general adaptive design and then the limit behavior of maximum quasi-likelihood estimators for a generalized linear model with an adaptive design. We will describe in this section how an adaptive design can be constructed and then we will look at some particular cases.

Recall the main features of the design:

- Patients arrive sequentially.
- Each patient can be assigned to either one of the two treatments and will be assigned to exactly one of them.
- The response from patients is observed immediately, prior to the assignment of the treatment to the next patient.

- The response is possibly dependent on a vector of covariates. The covariates are observed prior to the treatment assignment.
- We want the adaptive design to be such that the treatment allocation depends on the previous responses and the current covariate of the patient.

We will keep the same notation as in the previous chapters. Denote by A and B the two treatments that patients receive. Let t_k be the treatment allocation to the k^{th} patient, as defined by:

$$t_{k} = \begin{cases} 1 & \text{if } k^{th} \text{ patient is assigned to treatment A} \\ 0 & \text{if } k^{th} \text{ patient is assigned to treatment B} \end{cases}$$

Recall also that X_k denotes the *p*-dimensional covariate vector of the k^{th} patient and Y_k is the potential response of the k^{th} patient. We consider the generalized linear model introduced in Chapter 3 by relations (3.4) and (3.6). In particular $g(\cdot)$ is the link function to connect the mean of the response Y_k with the linear combination of the covariate X_k and treatment t_k .

We will use the notation Y_k^A and Y_k^B for the potential responses from treatment A and treatment B respectively and Y_k for the observed response:

$$Y_k = Y_k^A I[t_k = 1] + Y_k^B I[t_k = 0]$$
.

Assume, as before, the existence of an increasing sequence of σ -algebra $\{\mathcal{F}_k\}_{k\geq 1}$ such that the triplet (Y_k, X_k, t_k) is \mathcal{F}_k -measurable:

$$\mathcal{F}_{k} := \sigma(Y_{i}^{A}, Y_{i}^{B}, X_{i}, t_{i}, U_{i} : 1 \leq i \leq k) ,$$

where $(U_k)_{k\geq 1}$ denotes a sequence of i.i.d. random variables uniform distributed on (0, 1) and independent of the other variables in the model (Y_k^A, Y_k^B, X_k) . We will use the sequence $(U_k)_{k\geq 1}$ for randomizing the allocation treatment as in the construction given by (2.2).

Recall also the construction of the σ -algebras \mathcal{G}_k and that the treatment t_{k+1} is \mathcal{G}_k -measurable:

$$\mathcal{G}_k := \mathcal{F}_k \lor \sigma(X_{k+1}) \lor \sigma(U_{k+1})$$
.

In general we will say that we are at stage k if we have observed the first k-1 patients, and the covariate of the k^{th} patient, but we haven't observed yet the response from the k^{th} patient.

Let \hat{Y}_k^A and \hat{Y}_k^B be the predictors of the potential response of the k^{th} patient from treatment A, respectively treatment B. We want that allocation to treatment A to be made based on a function of these predictors, $f(\hat{Y}_k^A, \hat{Y}_k^B)$. We'll use the notation:

(4.1)
$$\hat{\pi}_k = f(\hat{Y}_k^A, \hat{Y}_k^B) \; .$$

It is also useful to think of $\hat{\pi}_k$ as an estimate of the parameter π_k , where π_k is the function of the unobserved potential responses:

$$\pi_k = f(Y^A_k, Y^B_k)$$
.

In summary, we consider the following general framework for our adaptive design:

• First stage

Start with n_0 patients allocated to each treatment.

• Second stage

At each stage $k > 2n_0$ we iterate the following procedure:

- Compute the maximum quasi-likelihood estimators $\hat{\beta}_{k-1,1}$, $\hat{\beta}_{k-1,2}$, $\hat{\beta}_{k-1,3}$ based on the information from the first k-1 patients as defined by relation (3.9). Recall that $\hat{\beta}_{k-1,1}$, $\hat{\beta}_{k-1,3}$ are p-dimensional column vectors and $\hat{\beta}_{k-1,2}$ is scalar.
- Observe the covariate vector X_k for the k^{th} patient.
- Compute the predictors \hat{Y}^A_k and \hat{Y}^B_k for the responses:

(4.2)
$$\hat{Y}_{k}^{A} = g(X_{k}\hat{\beta}_{k-1,1} + \hat{\beta}_{k-1,2} \cdot 1 + X_{k}\hat{\beta}_{k-1,3} \cdot 1)$$
$$= g(X_{k}(\hat{\beta}_{k-1,1} + \hat{\beta}_{k-1,3}) + \hat{\beta}_{k-1,2})$$

(4.3)
$$\hat{Y}_{k}^{B} = g(X_{k}\hat{\beta}_{k-1,1} + \hat{\beta}_{k-1,2} \cdot 0 + X_{k}\hat{\beta}_{k-1,3} \cdot 0)$$
$$= g(X_{k}\hat{\beta}_{k-1,1}) .$$

- Evaluate the function $\hat{\pi}_{k} := f(\hat{Y}_{k}^{A}, \hat{Y}_{k}^{B}).$
- Generate the treatment allocation t_k according to a $Bernoulli(\hat{\pi}_k)$ distribution. To achieve this, consider $(U_i)_{i\geq 1}$ the sequence of i.i.d. uniformly distributed random variables on interval [0, 1], independent of $(X_k, Y_k)_{k\geq 1}$. Then let t_k to be defined as:

$$(4.4) t_k := I[U_k < \hat{\pi}_k]$$

where the function $I[\cdot]$ is the usual indicator function.

Note that based on the proposed model the estimators \hat{Y}_k^A , \hat{Y}_k^B are computed from the information up to the $(k-1)^{th}$ patient and the current covariate X_k . Thus the treatment allocation t_k , through the dependence on the $\hat{\pi}_k$, involves all the information available at stage k-1 which is exactly our proposed goal stated in the beginning of the chapter.

The first stage of the design is needed for an initial estimation of the parameters of the model. If some estimators are already available, this first stage can be skipped and the design can start directly from the second stage.

From an ethical point of view the choice of the function $f(\cdot, \cdot)$ should satisfy some requirements according to the basic principle that we should favor the treatment with a superior response in the trial up to date.

As an example, in the case of a continuous response where larger values corresponding to better treatment, we propose the following guiding rules for constructing the function f:

(4.5)
$$f(x,x) = 0.5$$

(4.6)
$$\lim_{\substack{x \\ y \to \infty}} f(x, y) = 1$$

(4.7)
$$\lim_{\substack{y \\ x \to \infty}} f(x, y) = 0$$

Equation (4.5) expresses that in the case of same potential responses from the two treatments, we should allocate with probability 0.5. Equation (4.6) reflects that in the case of a superior potential response for treatment A then we should allocate to treatment A with a probability close to 1. Likewise, in the case (4.7), when the superior potential response comes from treatment B, then we should allocate to treatment A with a probability close to 0.

4.2 A Normal Response Example

In this section we investigate an example from the family of designs proposed in the beginning of the chapter. For the model (3.4) introduced in Chapter 3, we will concentrate on the case when the response is normally distributed together with the identity(canonical) link function for the model. In this particular situation the generalized linear model reduces to the linear model and the maximum quasilikelihood estimators become the least squares estimators. Keep the same notation introduced in the beginning of the chapter and consider the following model for the response:

(4.8)
$$Y_k = \beta_1 + \beta_2 X_k + \beta_3 t_k + \beta_4 X_k t_k + \epsilon_k$$

with the errors ϵ_k independent and normally distributed:

(4.9)
$$\epsilon_k \sim N(0, \sigma^2) \,.$$

We consider the covariate X_k to be unidimensional. The random covariates $(X_k)_{k\geq 1}$ are an i.i.d. sequence whose distributions have a known density p(x). The treatment allocation t_k is as described in the general framework of an adaptive design (see page 46). More precisely, we will consider

Let \mathbf{Z}_n be the model (4.8) design matrix:

(4.10)
$$\mathbf{Z}_{n} = \begin{pmatrix} 1 & X_{1} & t_{1} & X_{1} \cdot t_{1} \\ 1 & X_{2} & t_{2} & X_{2} \cdot t_{2} \\ \vdots & & & \\ 1 & X_{n} & t_{n} & X_{n} \cdot t_{n} \end{pmatrix}$$

Consider now the case when larger values for the response Y_k are desirable. Let

 $\hat{\beta}_{1,k}, \hat{\beta}_{2,k}, \hat{\beta}_{3,k}, \hat{\beta}_{4,k}$ be the least squares estimators for the parameters $\beta_1, \beta_2, \beta_3, \beta_4$, respectively, after k patients are treated and observed in the trial. Then the predictors for the k^{th} patient response from treatment A, respectively B are:

$$\begin{split} \hat{Y}_{k}^{A} &= \hat{\beta}_{1,k-1} + \hat{\beta}_{3,k-1} + (\hat{\beta}_{2,k-1} + \hat{\beta}_{4,k-1}) X_{k} \\ \hat{Y}_{k}^{B} &= \hat{\beta}_{1,k-1} + \hat{\beta}_{2,k-1} X_{k} \; . \end{split}$$

From ethical reasons we want to allocate the patients to the better treatment so we will 'favor' the treatment with a larger value from the two predictors \hat{Y}_k^A , \hat{Y}_k^B . At the same time we want to keep the randomization for the trial design. Therefore 'favoring' the better treatment will consist of allocating with a higher probability to the better treatment.

We will consider the following criteria for evaluating the better treatment. Let Δ_k be the difference for the k^{th} patient between the mean responses to the two treatments:

(4.11)
$$\Delta_k = E(Y_k^A) - E(Y_k^B)$$
$$= \beta_3 + \beta_4 X_k$$

and let $\hat{\Delta}_k$ be its estimator at stage k:

(4.12)
$$\hat{\Delta}_{k} = \hat{Y}_{k}^{A} - \hat{Y}_{k}^{B}$$
$$= \hat{\beta}_{3,k-1} + \hat{\beta}_{4,k-1} X_{k} .$$

A positive Δ_k shows that treatment A is better and analogously, if Δ_k is negative then the treatment B is better. Moreover a small absolute value of Δ_k is an

indication that the two treatments are not very different while a large absolute value of Δ_k is an indication that one of the treatments is much better than the other one.

Using the above interpretations we propose the following function for $\hat{\pi}_k$:

(4.13)
$$\hat{\pi}_k = \frac{\exp(\hat{\Delta}_k)}{1 + \exp(\hat{\Delta}_k)}$$

To emphasize that the treatment allocation is a function of \hat{Y}^A_k, \hat{Y}^B_k rewrite $\hat{\pi}_k$ as:

$$\hat{\pi}_{k} = \frac{\exp(\hat{\beta}_{3,k-1} + \hat{\beta}_{4,k-1}X_{k})}{1 + \exp(\hat{\beta}_{3,k-1} + \hat{\beta}_{4,k-1}X_{k})}$$
$$= \frac{\exp(\hat{Y}_{k}^{A} - \hat{Y}_{k}^{B})}{1 + \exp(\hat{Y}_{k}^{A} - \hat{Y}_{k}^{B})}.$$

Since in this case we will work with the estimator $\hat{\Delta}_k$, is easier to think of the probability of selection $\hat{\pi}_k$ as a function of it. Therefore we will refer in this case to $\hat{\pi}_k$ as a function of just one variable instead of two, $\hat{\pi}_k = f_1(\hat{\Delta}_k)$.

It is easy to observe that the three properties of the function $f(\cdot)$ as they were defined in (4.5) - (4.7) become the following properties for function $f_1(\cdot)$:

$$f_1(0) = 0.5$$
$$\lim_{x \to \infty} f_1(x) = \lim_{x \to \infty} \frac{\exp(x)}{1 + \exp(x)} = 1$$
$$\lim_{x \to -\infty} f_1(x) = \lim_{x \to -\infty} \frac{\exp(x)}{1 + \exp(x)} = 0.$$

In our case, $f_1(y) = \frac{\exp(y)}{1 + \exp(y)}$ and the three properties from above are satisfied, as it can be seen from the graph shown in figure 4.1.

The design is now completely specified according to the general framework presented in the introduction of the chapter (see page 47)



Figure 4.1: Graph of function $f_1(y) = \frac{\exp(y)}{1 + \exp(y)}$

4.2.1 An Evaluation Function for the Number of Patients on the Inferior Treatment

We will consider now an evaluation criterion for the design proposed in Chapter 4.1 from an ethical point of view. Recall that one of the main advantages of the adaptive designs is that they offer the choice of combining ethical considerations with the randomization needed for an impartial trial. In our case we want to allocate as many patients as we can to the better treatment, where the better treatment is decided based on the estimators from the model (4.8), but we have to keep in mind that the allocation is also subject to randomization introduced by (4.4).

The evaluation criterion is a natural one, given the model and the design considered. We will count the number of *'mistreatments'*, i.e. the number of allocations to the inferior treatment. It can be regarded as a *loss function*. We denote by C_k the number of 'mistreatments' up to stage k and is defined as follows:

(4.14)
$$C_k = \sum_{i=2n_0+1}^k I[t_i = 1] \cdot I[\Delta_i \le 0] + \sum_{i=2n_0+1}^k I[t_i = 0] \cdot I[\Delta_i > 0]$$

where we denoted by $I(\cdot)$ the usual indicator function.

The sum begins from $2n_0 + 1$ because we will take in consideration only the allocations done according to the design as defined by relations (4.13) and (4.12) and will not include the first $2n_0$ allocations from the first stage.

Observe that the variation in the loss function C_k comes from two sources: one is the variation coming from the randomization of the treatment assignment according to a Bernoulli distribution and the other comes from the parameter estimation we use in the treatment assignment.

We are interested in evaluating C_k . Based on the theorems we proved in the Chapter 3.3 we will study the behavior of C_k as $k \to \infty$ and we will be able to find a limit for the proportion of these number of patients allocated to the inferior treatment.

Theorem 4.2.1. Assume that the density function p(x) is uniformly bounded, the errors ϵ_k satisfy condition (3.14), the minimum and maximum eigenvalue of the design matrix \mathbf{Z}_n satisfy condition (3.13) and that

$$(4.15) \qquad \qquad \sup_{i>1} |X_i| < \infty \ a.s.$$

Then

$$(4.16) \qquad \qquad \frac{C_k}{k} \xrightarrow[k \to \infty]{} \int_{\beta_3 + \beta_4 x < 0} f_1(\beta_3 + \beta_4 x) p(x) \, dx + \int_{\beta_3 + \beta_4 x > 0} \left[1 - f_1(\beta_3 + \beta_4 x) \right] p(x) \, dx \, .$$

Proof. We begin by checking the assumptions of theorem (3.3.1) from chapter 3.

- Since in our case the model considered (4.8) is the linear model, the link function is the identity function, g(y) = y, which is continuously differentiable with positive derivative function, and therefore the first condition (3.11) from theorem (4.2.1) is satisfied.
- In our case we have the intercept and a one dimensional covariate, X_k , so condition (3.12) becomes:

$$\sup_{i\geq 1}\|(1,X_i)\|_2<\infty$$

But

$$||(1, X_i)||_2 = \sqrt{1 + X_i^2} < 1 + |X_i|$$

and using condition (4.15) the conclusion follows.

• Finally, conditions (3.13) and (3.14) follow from the assumptions made in the theorem for our model.

Therefore we can apply theorem (3.3.1) and conclude that:

$$(\hat{\beta}_{1,k}, \hat{\beta}_{2,k}, \hat{\beta}_{3,k}, \hat{\beta}_{4,k}) \rightarrow (\beta_1, \beta_2, \beta_3, \beta_4)$$
 a.s.

Recall the definition of the σ -algebra \mathcal{J}_k as defined by (2.5):

$$\mathcal{J}_k := \mathcal{F}_k \vee \sigma(X_{k+1})$$

and the equation (2.6):

$$E(t_{k+1}|\mathcal{J}_k) = \hat{\pi}_{k+1}$$

From the construction of the σ -algebra, X_k is \mathcal{J}_{k-1} -measurable. Since in our case

 $\Delta_k = \beta_3 + \beta_4 X_k$ it follows that Δ_k is also \mathcal{J}_{k-1} -measurable. Then for any $i > 2n_0+1$:

$$\begin{split} E(I[t_i = 1] \cdot I[\Delta_i \le 0]) &= E(E(I[t_i = 1] \cdot I[\Delta_i \le 0] | \mathcal{J}_{i-1}))) \\ &= E(I[\Delta_i \le 0] \cdot E(I[t_i = 1] | \mathcal{J}_{i-1}))) \\ &= E(I[\Delta_i \le 0] \cdot E(t_i | \mathcal{J}_{i-1})) \\ &= E(\hat{\pi}_i \cdot I[\Delta_i \le 0]) \end{split}$$

Analogous,

$$E(I[t_i = 0] \cdot I[\Delta_i > 0]) = E((1 - \hat{\pi}_i) \cdot I[\Delta_i > 0])$$

Then we have that

$$E\left(\frac{C_{k}}{k}\right) = \frac{1}{k} \sum_{i=2n_{0}+1}^{k} E(I[t_{i}=1] \cdot I[\Delta_{i} \leq 0]) + \frac{1}{k} \sum_{i=2n_{0}+1}^{k} E(I[t_{i}=0] \cdot I[\Delta_{i} > 0]) \\ = \frac{1}{k} \sum_{i=2n_{0}+1}^{k} E(\hat{\pi}_{i} \cdot I[\beta_{3} + \beta_{4}X_{i} \leq 0]) + \frac{1}{k} \sum_{i=2n_{0}+1}^{k} E((1 - \hat{\pi}_{i}) \cdot I[\beta_{3} + \beta_{4}X_{i} > 0]) \\ = \frac{1}{k} \sum_{i=2n_{0}+1}^{k} \int_{\beta_{3}+\beta_{4}x \leq 0} f_{1}(\hat{\beta}_{3,i-1} + \hat{\beta}_{4,i-1}x)p(x) \, dx + \frac{1}{k} \sum_{i=2n_{0}+1}^{k} \int_{\beta_{3}+\beta_{4}x \leq 0} (1 - f_{1}(\hat{\beta}_{3,i-1} + \hat{\beta}_{4,i-1}x))p(x) \, dx \\ = \frac{k - 2n_{0}}{k} \int_{\beta_{3}+\beta_{4}x \leq 0} \frac{1}{k - 2n_{0}} \sum_{i=2n_{0}+1}^{k} (1 - f_{1}(\hat{\beta}_{3,i-1} + \hat{\beta}_{4,i-1}x)p(x) \, dx + \frac{k - 2n_{0}}{k} \int_{\beta_{3}+\beta_{4}x \geq 0} \frac{1}{k - 2n_{0}} \sum_{i=2n_{0}+1}^{k} (1 - f_{1}(\hat{\beta}_{3,i-1} + \hat{\beta}_{4,i-1}x))p(x) \, dx$$

But from our previous statement, we have that $\hat{\beta}_{3,k} \to \beta_3$ and $\hat{\beta}_{4,k} \to \beta_4$ a.s. and because the function $f_1(\cdot)$ is continuous, we have that:

$$f_1(\hat{eta}_{3,i-1}+\hat{eta}_{4,i-1}x)\stackrel{a.s.}{
ightarrow} f_1(eta_3+eta_4x) \quad orall x$$

And therefore making use of the Stolz-Cesaro theorem (see A.1.1) we have that the series is convergent:

$$\frac{1}{k-2n_0} \sum_{i=2n_0+1}^k f_1(\hat{\beta}_{3,i-1} + \hat{\beta}_{4,i-1}x) p(x) \xrightarrow{a.s.} f_1(\beta_3 + \beta_4 x) p(x) \quad \forall x .$$

In the similar way we argue to get that:

$$\frac{1}{k-2n_0}\sum_{i=2n_0+1}^k (1-f_1(\hat{\beta}_{3,i-1}+\hat{\beta}_{4,i-1}x))p(x) \xrightarrow{a.s.} (1-f_1(\beta_3+\beta_4x))p(x) \quad \forall x \; .$$

But the function $f_1(\cdot)$ is uniformly bounded and from assumptions the density $p(\cdot)$ is also uniformly bounded, so

$$\sup_{x} |f_1(x)| \lor |p(x)| \le B$$

where B is some constant.

It follows from the dominated convergence theorem that:

$$\begin{split} \int_{\beta_{3}+\beta_{4}x\leq0} \frac{1}{k-2n_{0}} \sum_{i=2n_{0}+1}^{k} f_{1}(\hat{\beta}_{3,i-1}+\hat{\beta}_{4,i-1}x)p(x) \, dx \\ \xrightarrow{a.s.}{\longrightarrow} \int_{\beta_{3}+\beta_{4}x\leq0} f_{1}(\beta_{3}+\beta_{4}x)p(x) \, dx, \\ \int_{\beta_{3}+\beta_{4}x>0} \frac{1}{k-2n_{0}} \sum_{i=2n_{0}+1}^{k} (1-f_{1}(\hat{\beta}_{3,i-1}+\hat{\beta}_{4,i-1}x))p(x) \, dx \\ \xrightarrow{a.s.}{\longrightarrow} \int_{\beta_{3}+\beta_{4}x>0} (1-f_{1}(\beta_{3}+\beta_{4}x))p(x) \, dx \, . \end{split}$$

Combining the two relations together with (4.17) and using that $\frac{k-2n_0}{k} \to 1$ as $k \to \infty$ we get the conclusion of the theorem.

Let look more closely at the previous result in the particular case when X_k is uniformly distributed:

$$X_{k} \sim U(a, b)$$
.

Let L be the limit of the $\frac{C_k}{k}$ as it was defined by equation (4.16). Then we have the following corollary:

Corollary 4.2.2. If the covariate X_k from the model introduced by (4.8) and (4.9) is uniformly distributed on interval [a, b] then the limit L is equal to:

• Case 1: $a < -\frac{\beta_3}{\beta_4} < b$ (4.18) $L = \frac{1}{|\beta_4|(b-a)} \left(\ln \frac{4}{(1+e^{\beta_3+\beta_4 a})(1+e^{\beta_3+\beta_4 b})} + (\beta_3+\beta_4 a) \lor (\beta_3+\beta_4 b) \right)$

• Case 2:
$$-\frac{\beta_3}{\beta_4} < a$$

(4.19)
$$L = 1 - \frac{1}{|\beta_4|(b-a)} \ln \frac{1 + e^{\beta_3 + \beta_4 a}}{1 + e^{\beta_3 + \beta_4 b}}$$

• Case 3:
$$b < -\frac{\beta_3}{\beta_4}$$

(4.20)
$$L = \frac{1}{|\beta_4|(b-a)} \ln \frac{1 + e^{\beta_3 + \beta_4 a}}{1 + e^{\beta_3 + \beta_4 b}}$$

• Case 4:
$$\beta_4 = 0$$

(4.21)
$$L = \frac{e^{\beta_3}}{1 + e^{\beta_3}} I[\beta_3 \le 0] + \frac{1}{1 + e^{\beta_3}} I[\beta_3 > 0]$$

Proof. In this particular case the density function p(x) is equal to

$$p(x) = \frac{1}{b-a}I[a \le x \le b]$$

and the function $f_1(\cdot)$ was defined as $f_1(y) = \frac{e^y}{1+e^y}$. So the limit L given by rela-

tion (4.16) becomes:

$$\begin{split} L = & \int_{\{\beta_3 + \beta_4 x \le 0\} \cap \{a < x < b\}} \frac{1}{b-a} \cdot \frac{e^{\beta_3 + \beta_4 x}}{1 + e^{\beta_3 + \beta_4 x}} \, dx + \\ & \int_{\{\beta_3 + \beta_4 x > 0\} \cap \{a < x < b\}} \frac{1}{b-a} \cdot \left(1 - \frac{e^{\beta_3 + \beta_4 x}}{1 + e^{\beta_3 + \beta_4 x}}\right) \, dx \; . \end{split}$$

Calculus computations finish the proof.

As it can be seen from the above result we can distinguish there different scenarios for the example considered. The regression line for the response from treatment A and the regression line for the response from treatment B may intersect or not and if they do intersect then the point of intersection may be or not in the interval [a, b]. The case when they intersect inside the interval [a, b] corresponds to case 1, when they intersect to the left or to the right of interval [a, b] corresponds to case 2, respectively case 3 and finally when they are parallel corresponds to case 4. The four situations are illustrated in figure 4.2.

Figure 4.2: Illustration of Corollary 4.2.2



4.2.2 Simulations of the Design

In this section we will present some numerical results obtained from simulations. We will compute by Monte-Carlo methods the estimators for the coefficient parameters $(\beta_1, \beta_2, \beta_3, \beta_4)$ of model (4.8) and also we will compute the number of allocations to the inferior treatment, C_n , and the proportion $C_n/(n-2n_0)$. Although the theorems presented in the previous sections are all proved in the asymptotic case we will see that even in the case of moderate sample sizes the asymptotic results hold.

We will consider the simulations in the case when covariate X_k is uniformly distributed, presented in corollary (4.2.2).

Consider the following numerical application:

- $\beta_1 = 0, \ \beta_2 = 1, \ \beta_3 = 3, \ \beta_4 = -0.5.$
- $\sigma = 0.1, 0.2, 0.5, 1 \text{ or } 2.$
- a=0, b=10
- the fixed sample size is n = 30, 50, 100, and $n_0 = 5$, so that the first five patients are allocated to treatment A, the next five to treatment B and the next n-10 allocations are done according to our design introduced in page 46

Because of choice of the parameters $\beta_1, \beta_2, \beta_3, \beta_4$ we are in case 1 of the above result (4.2.2) and we get that in this particular case L = 0.242. The two regression lines for treatment A, respectively B are graphed below in figure 4.3. We use a continuous line to denote the regression line for Y^A and a dashed line to denote the reggression line for Y^B :

$$E(Y_k^A) = 3 + 0.5X_k$$
$$E(Y_k^B) = X_k$$


For a better understanding of the allocation design it is worth mentioning how the allocation probabilities vary as function of covariate X. Recall that

$$\Delta_x = \beta_3 + \beta_4 x$$
$$\pi_x = f_1(\Delta_x) = \frac{\exp(\Delta_x)}{1 + \exp(\Delta_x)}$$

and so in our case

$$\pi_x = \frac{\exp(3 - 0.5x)}{1 + \exp(3 - 0.5x)}$$

Then on the interval [0,10] the allocation probabilities vary according to the following table 4.1:

x	Δ_x	$f_1(\Delta_x)$
0	3	0.95
1	2.5	0.92
2	2	0.88
3	1.5	0.82
4	1	0.73
5	0.5	0.62
6	0	0.5
7	-0.5	0.38
8	-1	0.27
9	-1.5	0.18
10	-2	0.12

Table 4.1: Allocation probabilities to treatment A

A typical simulation of the design allocation is given in the following figure 4.4. We use symbol 'o' for marking the allocations to treatment A and symbol '+' for marking the allocations to treatment B.

Figure 4.4: A simulation design



In the following tables 4.2 - 4.7 we used Monte Carlo simulations to evaluate the bias and the variance of the estimators $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ and also to evaluate the estimate and the standard deviation for C_n and $C_n/(n - 2n_0)$. Since there are ten observations in the first stage, $n_0 = 5$, then for n = 100, there are ninety observations in the second stage, $n - n_0 = 100 - 2 \cdot 5 = 90$. In the limit L = 0.242 so we may expect to observe about $90 \cdot 0.242 = 21.78$ 'mistreatments' for C_{100} . Analogues, we may expect $40 \cdot 0.242 = 9.68$ 'mistreatments' for C_{50} and $20 \cdot 0.242 = 4.84$ 'mistreatments' for C_{30} . The estimates from the simulations were close to these values as it can be seen from the tables below.

In each of the following cases there are 10000 replications of the simulation of the allocation designs.

In figures 4.5 - 4.9 we can see the histogram from the 10000 replications of the simulation for the bias of the estimators $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ and also the histogram for the estimates of C_{100} . The histograms are shown in the case when n = 100, and $\sigma = 0.1, 0.5, 1$ and 2.

		bias	standard deviation
β_1			
	$\sigma = 0.1$	-0.0007	0.0762
	$\sigma = 0.2$	-0.0025	0.1523
	$\sigma = 0.5$	-0.0239	0.3870
	$\sigma = 1$	-0.0798	0.7851
	$\sigma = 2$	-0.2787	2.3751
β_2			
	$\sigma = 0.1$	0.0001	0.0100
	$\sigma {=} 0.2$	0.0002	0.0224
	$\sigma = 0.5$	0.0014	0.0583
	$\sigma = 1$	0.0047	0.1204
	$\sigma = 2$	0.0129	0.2590
β_3			
	$\sigma = 0.1$	0.0010	0.0900
	$\sigma = 0.2$	0.0012	0.1780
	$\sigma = 0.5$	0.0253	0.4629
	$\sigma = 1$	0.0861	0.9302
	$\sigma = 2$	0.2092	2.0170
eta_4			
	$\sigma = 0.1$	-0.0003	0.0141
	$\sigma = 0.2$	-0.0002	0.0300
	$\sigma {=} 0.5$	-0.0039	0.0755
	$\sigma = 1$	-0.0144	0.1597
	$\sigma=2$	-0.0330	0.3524

Table 4.2: Simulation results for $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$, in the case n = 30

		bias	standard deviation
β_1			
	$\sigma = 0.1$	-0.0021	0.0632
	$\sigma{=}0.2$	-0.0041	0.1261
	$\sigma = 0.5$	-0.0261	0.3122
	$\sigma = 1$	-0.0883	0.6534
	$\sigma = 2$	-0.3109	1.3500
β_2			
	$\sigma{=}0.1$	0.0002	0.0100
	$\sigma = 0.2$	0.0004	0.0173
	$\sigma = 0.5$	0.0027	0.0447
	$\sigma = 1$	0.0070	0.1170
	$\sigma = 2$	0.0212	0.2245
β_3			
	$\sigma = 0.1$	0.0021	0.0728
	$\sigma = 0.2$	0.0043	0.1428
	$\sigma {=} 0.5$	0.0287	0.3585
	$\sigma = 1$	0.0997	0.8209
	$\sigma{=}2$	0.2973	1.6500
β_4			
	$\sigma = 0.1$	-0.0003	0.0100
	$\sigma = 0.2$	-0.0008	0.0224
	$\sigma = 0.5$	-0.0048	0.0583
	$\sigma = 1$	-0.0160	0.1453
	$\sigma = 2$	-0.0478	0.2900

Table 4.3: Simulation results for $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$, in the case n = 50

		bias	standard deviation
β_1			
	$\sigma = 0.1$	-0.0009	0.0469
	$\sigma = 0.2$	-0.0035	0.0949
	$\sigma{=}0.5$	-0.0197	0.2328
	$\sigma = 1$	-0.0644	0.4824
	$\sigma=2$	-0.2559	1.0516
β_2			
	$\sigma = 0.1$	0.0001	0.0100
	$\sigma = 0.2$	0.0004	0.0141
	$\sigma {=} 0.5$	0.0021	0.0332
	$\sigma = 1$	0.0067	0.0678
	$\sigma = 2$	0.0216	0.1539
β_3			
	$\sigma = 0.1$	0.0010	0.0529
	$\sigma = 0.2$	0.0044	0.1068
	$\sigma {=} 0.5$	0.0217	0.2615
	$\sigma = 1$	0.0713	0.5530
	$\sigma{=}2$	0.2711	1.2147
β_4			
	$\sigma{=}0.1$	-0.0002	0.0100
	$\sigma = 0.2$	-0.0007	0.0173
	$\sigma{=}0.5$	-0.0034	0.0424
	$\sigma = 1$	-0.0116	0.0877
	$\sigma = 2$	-0.0416	0.1977

Table 4.4: Simulation results for $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$, in the case n = 100

		estimate	standard deviation
C_{30}			
	$\sigma = 0.1$	4.8632	1.8988
	$\sigma = 0.2$	4.8698	1.9196
	$\sigma = 0.5$	4.9634	2.0262
	$\sigma = 1$	5.2542	2.2815
	$\sigma = 2$	5.9658	3.0394
$C_{30}/20$			
	$\sigma = 0.1$	0.2432	0.0949
	$\sigma = 0.2$	0.2435	0.0960
	$\sigma = 0.5$	0.2482	0.1013
	$\sigma = 1$	0.2627	0.1141
	$\sigma = 2$	0.2983	0.1520

Table 4.5: Simulation results for C_{30} , in the case n = 30

<u>.</u>		estimate	standard deviation
C_{50}			
	$\sigma = 0.1$	9.7533	2.7314
	$\sigma = 0.2$	9.6947	2.7323
	$\sigma = 0.5$	9.7995	2.8569
	$\sigma = 1$	10.1507	3.3095
	$\sigma = 2$	11.3508	4.8413
$C_{50}/40$			
	$\sigma = 0.1$	0.2438	0.0683
	$\sigma = 0.2$	0.2424	0.0684
	$\sigma = 0.5$	0.2450	0.0714
	$\sigma = 1$	0.2538	0.0827
	$\sigma = 2$	0.2838	0.1210

Table 4.6: Simulation results for C_{50} , in the case n = 50

		estimate	standard deviation
C_{100}			
	$\sigma = 0.1$	21.7849	4.0953
	$\sigma = 0.2$	21.8040	4.1247
	$\sigma {=} 0.5$	21.9066	4.3555
	$\sigma = 1$	22.3827	5.1867
	$\sigma=2$	24.0572	7.7539
$C_{100}/90$			
	$\sigma = 0.1$	0.2421	0.0455
	$\sigma = 0.2$	0.2423	0.0458
	$\sigma = 0.5$	0.2434	0.0484
	$\sigma = 1$	0.2487	0.0576
	$\sigma = 2$	0.2673	0.0860

Table 4.7: Simulation results for C_{100} , in the case n = 100

Figure 4.5: Histogram of 10000 replications for bias of $\hat{\beta_1}$ when n= 100, $\sigma=0.1, 0.5, 1, 2$



Figure 4.6: Histogram of 10000 replications for bias of $\hat{\beta_2}$ when n= 100, $\sigma=0.1, 0.5, 1, 2$



Figure 4.7: Histogram of 10000 replications for bias of $\hat{\beta}_3$ when $n = 100, \sigma = 0.1, 0.5, 1, 2$



Figure 4.8: Histogram of 10000 replications for bias of $\hat{\beta}_4$ when n= 100, $\sigma=0.1, 0.5, 1, 2$



Figure 4.9: Histogram of 10000 replications for estimates of C_{100} when n = 100, $\sigma = 0.1, 0.5, 1, 2$



Finally, we want to compare from an ethical point of view our design (ADC) with other designs proposed in the literature. So we will compare the number of *'mistreatments'*, C_n , and the proportion $C_n/(n-2n_0)$ from two other designs:

- complete randomization (CR) design where each treatment has 50% chance of being assigned (the covariate is ignored)
- deterministic (D) design where allocation is done according to the larger mean response of the two treatments, that is

$$t_{k} = I[\bar{Y}_{k-1}^{A} > \bar{Y}_{k-1}^{B}]$$

The results from the simulations with n = 50,100 and $\sigma = 0.1, 0.2, 0.5, 1, 2$ are presented in tables 4.8 - 4.11. Recall that for our design in the first stage we allocate 5 patients to each treatment, so for comparison purposes we have to take 100-10=90, respectively 50-10=40 allocations for the other two designs enumerated before.

estimate				
		ADC	CR	D
	$\sigma=0.1$	9.75	19.99	17.70
	$\sigma = 0.2$	9.70	20.01	17.71
	$\sigma = 0.5$	9.80	20.01	17.62
	$\sigma = 1$	10.15	19.99	17.87
	$\sigma{=}2$	11.35	20.04	18.31
standard deviation				
de viditoit		ADC	CR	D
	$\sigma=0.1$	2.73	3.13	3.72
	$\sigma = 0.2$	2.73	3.22	3.73
	$\sigma = 0.5$	2.86	3.13	3.74
	$\sigma = 1$	3.31	3.17	3.93
	$\sigma{=}2$	4.84	3.18	4.18

Table 4.8: Simulation results for C_{50} , in the case n = 50

	ADC	CR	D
$\sigma=0.1$	0.244	0.500	0.443
$\sigma = 0.2$	0.243	0.500	0.443
$\sigma = 0.5$	0.245	0.500	0.441
$\sigma = 1$	0.254	0.500	0.447
$\sigma=2$	0.283	0.501	0.458
	ADC	CR	D
$\sigma=0.1$	0.068	0.078	0.093
$\sigma = 0.2$	0.068	0.081	0.093
$\sigma = 0.5$	0.072	0.078	0.094
$\sigma = 1$	0.083	0.079	0.098
	$ \begin{array}{c} \sigma = 0.1 \\ \sigma = 0.2 \\ \sigma = 0.5 \\ \sigma = 1 \\ \sigma = 2 \end{array} $ $ \begin{array}{c} \sigma = 0.1 \\ \sigma = 0.2 \\ \sigma = 0.5 \end{array} $	$\begin{array}{c c} & ADC \\ \hline \sigma = 0.1 & 0.244 \\ \sigma = 0.2 & 0.243 \\ \sigma = 0.5 & 0.245 \\ \sigma = 1 & 0.254 \\ \sigma = 2 & 0.283 \\ \hline \end{array}$	$\begin{array}{c ccccc} ADC & CR \\ \hline \sigma = 0.1 & 0.244 & 0.500 \\ \sigma = 0.2 & 0.243 & 0.500 \\ \sigma = 0.5 & 0.245 & 0.500 \\ \sigma = 1 & 0.254 & 0.500 \\ \sigma = 2 & 0.283 & 0.501 \\ \hline \\ \hline \end{array}$

Table 4.9: Simulation results for $C_{50}/40$, in the case n = 50

estimate				
		ADC	CR	D
	$\sigma = 0.1$	21.78	44.98	38.98
	$\sigma = 0.2$	21.80	45.09	39.02
	$\sigma = 0.5$	21.91	44.98	39.27
	$\sigma = 1$	22.38	44.96	39.58
	$\sigma=2$	24.06	44.96	40.84
standard				
deviation		ADC	CR	D
	$\sigma=0.1$	4.10	4.73	6.73
	$\sigma = 0.2$	4.12	4.69	6.75
	$\sigma = 0.5$	4.36	4.80	6.98
	$\sigma = 1$	5.19	4.74	7.31
	$\sigma = 2$	7.75	4.74	8.12

Table 4.10: Simulation results for C_{100} , in the case n = 100

estimate				
		ADC	CR	D
	$\sigma = 0.1$	0.242	0.500	0.433
	$\sigma = 0.2$	0.242	0.501	0.434
	$\sigma = 0.5$	0.243	0.500	0.436
	$\sigma = 1$	0.249	0.499	0.440
	$\sigma=2$	0.267	0.499	0.454
standard				
deviation			<u> </u>	-
		ADC	CR	D
	$\sigma = 0.1$	0.046	0.053	0.075
	$\sigma = 0.2$	0.046	0.052	0.075
	$\sigma = 0.5$	0.048	0.053	0.078
	$\sigma = 1$	0.058	0.053	0.081
	$\sigma {=} 2$	0.086	0.053	0.090

Table 4.11: Simulation results for $C_{100}/90$, in the case n = 100

4.3 Conclusions and Future Work

The design we propose is quite natural and fits very well in situations described by the outline below:

- Delay in response is moderate, allowing the adaptation to take place.
- The recruitment can take place during most or all of the trial.
- It is known that a generalized linear model is a good fit for the relationship between the response, the covariates and the treatment.
- Modest gain in terms of treatment successes is desirable from an ethical standpoint.

The design can be especially useful if it is desirable also to have an estimation of the parameters of the model during the duration of the clinical trial before waiting for the entire data to be collected.

We have seen that the asymptotic results obtained in theorems 3.3.1 and 4.2.1 are consistent with the simulations even in the case when the n is relative small (n = 30, 50, 100). The histograms also show a normal behavior for the estimators, but as dispersion parameter σ^2 increases the variance of the estimators increase too, leading in some cases to more biased estimates.

There are several lines of research arising from the current thesis work which should be pursued.

- There is a natural extension to the case when in the clinical trial there are more than two treatments to be compared.
- One of the requirement of the design studied was to have responses available prior to the assignment of the treatment to the next patient. But as often is the case in practice, delayed responses should be considered in the design, too.

- Of particular interest is the extension of the normal asymptotic result from the normal linear model to the generalized linear model.
- An useful design is such that the function π_k that we use in randomizing the allocation is obtained from various optimal criteria. Of interest in this case is also to compare the optimal target with its correspondent from the adaptive design.
- Lot of work has been done in the literature in the case where the covariates are categorical, so a stratified design can be employed. An application of our design in this particular case would be interesting and the performance of the designs should compared.

Appendix A

A.1

Theorem A.1.1 (Stolz-Cesaro Theorem (Sireţchi (1985)). Let $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ be two sequences of real numbers. If b_n is positive, strictly increasing and unbounded and the following limit exists:

$$\lim_{n \to \infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n} = l$$

Then the limit:

$$\lim_{n\to\infty}\frac{a_n}{b_n}$$

also exists and it is equal to l.

Proof. From the definition of convergence, for every $\epsilon > 0$ there is $N(\epsilon) \in \mathbb{N}$ such that $(\forall)n \ge N(\epsilon)$, we have:

$$l - \epsilon < \frac{a_{n+1} - a_n}{b_{n+1} - b_n} < l + \epsilon$$

Because b_n is strictly increasing we can multiply the last equation with $b_{n+1} - b_n$ to get :

$$(l-\epsilon)(b_{n+1}-b_n) < a_{n+1}-a_n < (l+\epsilon)(b_{n+1}-b_n)$$

Let $k>N(\epsilon)$ be a natural number . Summing the last relation we get :

$$(l-\epsilon)\sum_{i=N(\epsilon)}^{k} (b_{i+1}-b_i) < \sum_{i=N(\epsilon)}^{k} (a_{n+1}-a_n) < (l+\epsilon)\sum_{i=N(\epsilon)}^{k} (b_{i+1}-b_i) \Rightarrow$$
$$(l-\epsilon)(b_{k+1}-b_{N(\epsilon)}) < a_{k+1}-a_{N(\epsilon)} < (l+\epsilon)(b_{k+1}-b_{N(\epsilon)})$$

Divide the last relation by $b_{k+1}>0$ to get :

$$(l-\epsilon)(1-\frac{b_{N(\epsilon)}}{b_{k+1}}) < \frac{a_{k+1}}{b_{k+1}} - \frac{a_{N(\epsilon)}}{b_{k+1}} < (l+\epsilon)(1-\frac{b_{N(\epsilon)}}{b_{k+1}}) \Leftrightarrow$$
$$(l-\epsilon)(1-\frac{b_{N(\epsilon)}}{b_{k+1}}) + \frac{a_{N(\epsilon)}}{b_{k+1}} < \frac{a_{k+1}}{b_{k+1}} < (l+\epsilon)(1-\frac{b_{N(\epsilon)}}{b_{k+1}}) + \frac{a_{N(\epsilon)}}{b_{k+1}}$$

This means that there is some K such that for $k \geq K$ we have :

$$(l-\epsilon) < \frac{a_{k+1}}{b_{k+1}} < (l+\epsilon)$$

since the other terms who were left out converge to 0.

This means that :

$$\lim_{n\to\infty}\frac{a_n}{b_n}=l\;.$$

Bibliography

- [1] Andersen, T.W. and Taylor, J. (1979). Strong consistency of least squares estimators in dynamic models, *The Annals of Statistics*, 7, 484-489.
- [2] Atkinson, A.C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors, *Biometrika*, 69, 61-67.
- [3] Blackwell, D. and Hodges, J.L. (1957). Design for the control of selection bias, Annals of Mathematical Statistics, 28, 446-460.
- [4] Chen, K., Hu, I. and Ying, Z. (1999). Strong consistency of maximum quasilikelihood estimators in generalized linear models with fixed and adaptive designs, *The Annals of Statistics*, 27, 1155-1163.
- [5] Clayton, M.K. (1989). Covariate models for Bernoulli bandits, Sequential Analysis, 8, 406-426.
- [6] Cornell, R.G., Landenberger B.D., Bartlett and R.H. (1986). Randomized play the winner clinical trials, Communications in Statistics-Theory and Methods, 15, 159-178.
- [7] Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 2, 513-535.
- [8] Efron, B. (1971). Forcing a sequential experiment to be balanced, *Biometrika*, 58, 403-417
- [9] Geraldes, M.C. (1999). Covariates in adaptive designs for Clinical Trials, Michigan State University Graduate School(doctoral dissertation)
- [10] Ivanova, A., Rosenberger, W., Durham, S. and Flournoy, N. (2000). A birth and death urn for randomized clinical trials: asymptotic methods Sankhyā: The Indian Journal of Statistics B, 62, 104-118.
- [11] Jenison, C. and Turnbull, B.W. (2000). Group Sequential Methods with Applications to Clinical Trials, Chapman and Hall, Boca Raton.

- [12] Lachin, J.M. (1988). Properties of simple randomization clinical trials Controlled Clinical Trials, 9, 312-326.
- [13] Lai, T.L. and Robbins, H. (1979). Adaptive design and stochastic approximation, The Annals of Statistics, 7, 1196-1221.
- [14] Lai, T.L., Robbins, H. and Wei, C.Z. (1979). Strong consistency of least squares estimates in multiple regression II, *Journal of Multivariate Analysis*, 9, 343-361.
- [15] Lai, T.L. and Wei, C.Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *The Annals of Statistics*, 10, 154-166.
- [16] Lai, T.L. and Robbins, H. (1982). Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 56, 329-360.
- [17] McCullagh, P. and Nelder, J.A. (1989). Generalized Linear Models 2nd ed., Chapman and Hall, London.
- [18] Melfi, V. and Page, C. (2000). Estimation after adaptive allocation, Journal of Statistical Planning and Inference, 87, 353-363.
- [19] Melfi, V., Page, C. and Geraldes, M. (2001). An adaptive randomized design with application to estimation, *Canadian Journal of Statistics*, 29, 107-116.
- [20] Pocock, S.J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognosis factors in the controlled clinical trial, *Biometrics*, 31, 103-115.
- [21] Robins, H. and Monro, S. (1951). A stochastic approximation method, Annals of Mathematical Statistics, 22, 400-407.
- [22] Rosenberger, W.F. (1999). Randomized play-the-winner clinical trials: review and recommendations, *Controlled Clinical Trials*, 20, 328-342.
- [23] Rosenberger, W. (2002). Randomized urn models and sequential design, Sequential Analysis, 21, 1-41.
- [24] Sarkar, J. (1991). One-armed bandit problems with covariates, The Annals of Statistics, 19, 1978-2002.
- [25] Sireţchi, G. (1985). Differential and Integral Calculus, vol.1, Editura Științifică și Enciclopedica, Bucharest
- [26] Smith, R.L. (1984). Sequential treatment allocation using biased coin designs, Journal of the Royal Statistics Society B, 46, 519-543.

- [27] Tamura, R.N., Faires D.E., Andersen J.S. and Heiligenstein, J.H. (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder, *Journal of the American Statistical Association*, 89, 768-776.
- [28] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method, *Biometrika*, 61, 439-447.
- [29] Wei, L.J. (1978). The adaptive biased coin design for sequential experiments, The Annals of Statistics, 6, 92-100.
- [30] Wei, L.J., Durham, S. (1978) The randomized play-the-winer rule in medical trials, Journal of the American Statistical Association, 73, 840-843.
- [31] Wu, C.F.J. (1985). Efficient sequential designs with binary data, Journal of American Statistical Association, 80, 974-984.
- [32] Yang, Y. and Zhu, D. (2002). Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates, *The Annals of Statistics*, 30, 100-121.