Design and inference of clinical trials with continuous covariates

Xiaoming Li

Linan, Zhejiang, P.R. China

M.S., Zhejiang University, P.R.China, 2011
B.S., East China Normal University, P.R.China, 2009

A Dissertation Presented to the Graduate Faculty
of University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Statistics

University of Virginia
December, 2015

Abstract

Covariate-adaptive designs are widely used to balance covariates and maintain randomization in clinical trials. Restricted randomization procedures for discrete covariates and their asymptotic properties have been addressed in the literature. However, clinical trials can often contain continuous covariates. Simply discretizing or categorizing continuous covariates can result in lose of information. The state-of-the-art adaptive design with continuous covariates is still entirely based on simulation and to-date lacks a rigorous theoretical understanding. Therefore, conventional hypothesis testing for clinical trials using continuous covariates is still not well understood. In this dissertation, we establish a theoretical framework for hypothesis testing on clinical trials with continuous covariates randomized using adaptive designs. We test for treatment effects and significance of covariates under null and alternative hypotheses. To verify our framework, numerical simulations are conducted under a class of covariate-adaptive designs including, the p-value based method, the Su's percentile method, the empirical cumulative-distribution method, the Kullback-Leibler divergence method, and the kernel-density method. For independent covariates we find that: (1) hypothesis testing that compares treatment effects has small Type I error, (2) hypothesis testing using adaptive designs outperforms complete randomization method in terms of power, and (3) testing for significance of covariates is still valid. For correlated covariates we prove and verify in simulations that treatment effects still have small Type I error, and estimators of continuous covariate coefficients are biased under covariate-adaptive designs. Furthermore, we adapt a minimization procedure to the kernel-density method for covariate-adaptive design, and show that our method out-performs other adaptive designs in balancing the distributions of continuous covariates across treatment groups in clinical trials.

Acknowledgements

I would like to sincerely thank my advisors, Prof. Feifang Hu and Prof. Jianhui Zhou, for guiding me through my Ph.D research and study at the University of Virginia. Both advisors were an integral part of my development and the development of this dissertation. Without their guidance and advice this dissertation would not be possible.

Both Prof. Hu and Prof. Zhou exuded great commitment for teaching and research. Their passion has inspired me to work hard and remain focused throughout my PhD candidacy. Also over the years, Prof. Feifang Hu has also been a life mentor, helping me through difficult times. I will always remember him telling his students to find their individual path and purpose in life. Moving forward, I will look for my path and achieve the best that I can in my career and life. Overall, I am grateful and humbled by the opportunity to have worked with Prof. Hu and Prof. Zhou. Collectively, they have pushed me to achieve more.

I would also like to extend my gratitude to committee members, Prof. Dan Keenan and Prof. Guofen Yan for their guidance and suggestions regarding my dissertation. I am also grateful to Prof. Tingting Zhang who taught me many statistics courses. Finally, I would like to thank Mrs. Karen Dalton for taking care of the day-to-day logistic in the department so that students can focus on learning and research.

A special thank goes to my research colleagues Yanqing Hu, Zhenjun Ma, Wei Ma, Yuan Xue and other classmates who helped me.

I owe my deepest gratitude to my family for their unconditional love and support throughout my life. My grandmother, my parents, my sister, and my husband have been my greatest support, and for that I am greatly indebted.

Contents

Chapter 1

Introduction

1.1 Overview of adaptive design

1.1.1 Background

In clinical trials, randomizing patient assignments and balancing patient allocation are integral for convincing treatment comparison. The aim of randomization is to reduce selection bias, however complete randomization methods, i.e., flip of an unbiased coin can be prone to imbalance especially in small clinical trials due to unpredictability. On the other hand, perfect balancing methods can make clinical trials susceptible to selection bias. However, in order to maximize the power of treatment comparison, we need to balance patient allocation without compromising randomization. To handle this dilemma, adaptive randomization or restricted randomization (Rosenberger and Lachin, 2002) designs have been introduced.

In this chapter, we review the four main classes of adaptive randomization. This includes restricted randomization, response-adaptive design, covariate-adaptive design and covariate-adjusted response adaptive design. With the focus of the disser-

tation on covariate-adaptive design, we review how to handle both discrete and continuous covariates. A review of hypothesis testing under covariate adaptive design is presented in the next section. In the final section we motivate this dissertation.

1.1.2 Four major classes of adaptive randomization

Adaptive design utilizes historical information from a clinical trial to adjust future treatment assignment. The information from a clinical trial includes: treatment assignments of previous patients, the covariates of the patients, and the treatment response of assigned patients. Based on the information used to assign new patients, the adaptive design is classified into four-main families (Hu and Rosenberger, 2006): (1) restricted randomization, (2) response-adaptive design, (3) covariate-adaptive design, and (4) covariate-adjusted-response adaptive design.

Restricted randomization is a procedure that uses past treatment assignments to select the probability of future treatment assignments, with the objective to balance the number of subjects across treatment groups. The most widely used restricted randomization procedure is the permuted block design (PBD). In the permuted block design, sample allocations are made at random within blocks of size $2m$, where m number of subjects are equally assigned to A and B treatments. In this case, there are $m, (m \geq 1)$ study subjects per trial. The PBD method perfectly balances allocation for at least two treatments. However, PBD method is susceptible to selection bias. Some allocations in the tail of each block can be guessed with high probability or even with certainty. Therefore, while the PBD method achieves the objective of balancing treatments, it fails to protect the study from experimental bias, due to its deterministic property. Berger et al. (2003) improved PBD by introducing a maximal procedure with the least restrictive allocation procedure

subject to a constraint on the maximum tolerated imbalance.

An important class of restricted randomization procedures is the biased coin designs (BCD). The first and the most widely used BCD was proposed by Efron (1971). In Efron's method, the difference between the number of patients in each treatment either skews the probability of assigning patients toward the underrepresented treatment or balances the trial with probability 1/2. Hu et al. (2009) show that Efron's BCD is an asymptotically best procedure in targeting balanced allocation with minimal variability for a two treatment trial. Since the 1970s, a variety of BCD methods have been developed and studied. These methods can be divided into two groups based on whether the procedures skew allocation probability according to the magnitude of the current treatment imbalance or simply use a fixed probability. The first group of biased-coin designs uses treatment difference as the measure of imbalance with fixed allocation probability. This includes works by Efron's (1971) BCD, Soares and Wu's (1983) big stick rule, Chen's (1999) BCD, and Baldi Antognini and Giovagnolli's (2004) adjustable BCD. The second group of BCDs skews the probability conditional on the current imbalance rather than using a fixed probability. This include works by Wei (1978), Atkinson (1982) and Smith's (1984) generalized biased coin designs (GBCDs).

Response-adaptive design utilizes patients' responses to their assigned treatments to adjust the allocation probability of new patients in order to achieve the desired allocation target (Hu and Rosenberger, 2003). The urn design is widely used in response-adaptive design. The idea behind the urn model was first proposed from an ethical point-of-view by assigning more patients to the better treatment. In the urn model, the assignment of the incoming patient is determined by the type of ball drawn from an urn containing two types of balls (red, black) representing two treatments, A and B, respectively. Wei and Durham (1978) introduced "randomized

play-the- winner" model, where a success adds more same type balls and failure adds more opposite type balls. Moreover, Ivanova (2003) proposed the "drop-the-loser" urn model, which replaces a ball for every success and drops it for every failure. The drop-the-loser rule is a fully randomized procedure with minimal variability (Rosenberger et al., 2012). The generalization of "drop-the-loser" urn model is found in Zhang et al. (2007), Sun et al. (2007) and Zhang, et al. (2011).

Modern research on response-adaptive design has focused on optimizing some specific criterion with chosen allocation target R. For example, consider a clinical trial with two treatments (A and B), the success rates $\theta = (p_A, p_B)$ and the failure rates (q_A, q_B). Neyman allocation: $R = \sqrt{p_A q_A}/\sqrt{p_B q_B}$, is designed to maximize the power under a fixed sample size. $R = \sqrt{p_A}/\sqrt{p_B}$ minimizes the expected number of failures for a fixed power (Rosenberger and Lachin, 2002). Even though these methods have attractive properties in theory, the actual success rates $\theta = (p_A, p_B)$ are unknown parameters. Thus the optimal allocation rate is unknown at the beginning of the trial, and needs to be estimated during the process, which leads to the doubly-adaptive biased coin design (DBCD) (Eisele, 1994, Eisele and Woodroofe, 1995). Hu and Zhang (2004) improved the DBCD method with conditions that are easy to verify. Their method is fully randomized with minimal variability and also efficiently converges to the chosen optimal target allocation. Two recent works regarding response-adaptive designs are the asymptotically best response-adaptive randomization procedures, Hu et al. (2006), and the efficient response adaptive randomization designs (ERADE), Hu et al. (2009).

Covariate-adaptive designs balance patients with respect to key covariates across treatment groups by changing the probability of patient assignment according to the previous treatment allocation and patients' covariate information. Real clinical trials require the balance of important baseline factors to get a convincing comparison of

the treatment effect. Baseline factors are well balanced by simple randomization when sample size is large enough. However, the balance of all key covariates is hard to achieve when sample size is small, which is common in real clinical trials. Though it is possible to remedy imbalance after the trial, we share the belief with many other researchers that balancing for prognostic factors prior to a trial is necessary, both in terms of efficiency and convincing (Begg 1980). Over the 50,000 trials conducted from 1989 to 2008 (Tave, 2010), more than 95% of the trials applied stratified permuted block design to balance covariates. More recently, minimization procedures (Tave, 1974; Pocock and Simon, 1975) have been gaining ground for use in clinical trials. More details of well-known covariate-adaptive designs will be discussed in section 1.2.

Covariate-adjusted response adaptive design (Zhang et al., 2007) is a procedure that randomizes according to an allocation function that depends on (1) all previous patients' covariates, (2) the coming patient's covariate vector and (3) all previous response. This approach is a personalized randomization procedure. It aims at finding a therapy or a dosage that is most appropriate for an individual patient, with the potential benefits of increasing efficacy as well as safety. Historical information of the covariates and the responses of previous patients are used to predict the responses of current patients to different treatments. The target allocation is a function of the response parameter and patients' covariate vectors. The first approach (Rosenberger et al., 2001) is to randomize patients with probabilities proportional to the current estimate of the treatment difference adjusting for patients' covariates. Zhang et al. (2007) proposed to determine the desired proportions of patients assignment and different covariate values based on specifying a target allocation function. Other works include weighted optimality approach by Atkinson (1982) and Atkinson and Biswas (2005), bayesian adaptive randomization methods

by Biswas and Angers (2002), Thall and Wathen (2005), and Cheung et al. (2006) and randomized survival methods by Zhu and Hu (2010).

1.2 Covariate-adaptive design

1.2.1 Handling discrete covariate in clinical trials

The balance of important baseline covariates, such as gender, age, disease stage, is essential in clinical trials. To maintain the validity of treatment comparisons, covariate adaptive allocation is often adopted in sequential clinical trials. Stratification is an intuitive method to balance categorical covariates. Strata are defined as the combinations of different levels of covariates. In handling discrete covariates, the most commonly used strategy is stratified permuted block design (SPBD). Stratified permuted block design based on discrete covariate is a well accepted method, where the permuted block allocation is employed within each strata. This method can achieve good balance for different covariates when the number of strata is small, but may cause selection bias for clinical trials with many strata. Especially with a moderate sample size and a large number of covariates, most strata have very few patients. In this case, SPB design performs as bad as complete randomization, whose marginal imbalance and overall imbalance can be extremely large.

To deal with the trials containing a large number of strata, minimization procedures (or named covariate-adaptive designs) are proposed in the mid-1970s. The first minimization procedure is proposed by Tave (1974). Tave's minimization method is a deterministic method which allocates patients to treatments to minimize imbalances on important covariates with probability 1. Pocock and Simon (1975) and Wei (1978) generalized minimization to randomized clinical trials. Pocock and Si-

mon's method considers a clinical trial with two treatments ($k = A, B$), and the balancing of covariates ($Z_1, \cdots, Z_I$) between these treatments. Suppose the $n+1$th patient to be randomized is a member of strata ($r_1, \cdots, r_I$) of covariates $Z_1, \cdots, Z_I$. Let $N_{ijk}(n), i = 1, \cdots, I, j = 1, \cdots, n_i, k = 1, 2, (1 = A, 2 = B)$, be the number of patients in stratum j of covariate i on treatment k after n patients have been randomized. Defining $D_i(n) = N_{ir_i1}(n) - N_{ir_i2}(n)$, the sum over weighted strata is $D(n) = \sum_{i=1}^{I} w_i D_i(n)$, where w_i are weights chosen based on importance of individual covariates. If $D(n)$ is less than 0, then the weighted measurement indicates that for set ($r_1, \cdots, r_I$) of strata and the new patient should be assigned with probability p ($p > \frac{1}{2}$) to treatment A; and vice-versa. If $D(n)$ is greater than 0, Pocock and Simon suggest assigning the next patient to treatment A with probability $1 - p$. If $D(n) = 1/2$ then the next patient is assigned to treatment A with probability 1/2. Pocock and Simon's minimization method controls both the overall difference across treatments as well as the marginal difference on important covariates.

Another important covariate-adaptive design is Wei's marginal procedure using urns. Wei (1977, 1978) developed an adaptive biased coin design where the probability of assigning patients adapts according to the degree of imbalance. At the beginning of the trial, an urn contains α balls of each of two types, A and B. Draw a ball and replace, if it is a type A ball, assign the patient to treatment A, and add β type B balls to the urn. When the number of covariates is large and the stratum sizes are small, use separate urn in each stratum. The one with the greatest imbalance is used to generate the patient assignment. Wei (1978) proves that if there is no interaction between the covariates or between the treatment effect and covariates in a standard linear model, then marginal balance is sufficient to achieve an unbiased estimate of the treatment difference. A review of urn model is found in Wei and Lachin (1988).

Many procedures that achieve balanced allocation for both margins of the covariates and within-strata are proposed by Zelen (1974), Nordle and Brandmark (1977), Efron (1980), Signorini, Leung et al. (1993), and Heritier, Gebski, et al. (2005). These methods take care of the balance within-stratum when interactions between the covariates exist.

Before Hu and Hu (2012), a theoretical understanding of minimization procedures for handling covariates was missing in the literature. For a limited family of covariate-adaptive designs based on bias coin allocation, Hu and Hu (2012) show that the imbalances are positive recurrent Markov chains. Later, Hu and Zhang (2013) proposed and determined the theoretical properties for an even more general family of discrete covariate-adaptive designs that incorporates most of the prior well-accepted discrete covariate-adaptive designs as special cases. They treat imbalance measure as a weighted average of overall difference, marginal difference and within-strata difference. For the case where only the overall imbalance measure is considered, Hu and Zhang's model reduces to Efron's biased coin design. If the weight on within-strata imbalance measures is zero, and only the marginal imbalances are considered, then Hu and Zhang's model reduces to Pocock and Simon's (1975) marginal method. Furthermore, if only the within-strata imbalance measure is considered, their model reduces to stratified randomization, where a separate biased coin is employed to determine the assignment within each stratum.

1.2.2 Handling continuous covariate in clinical trials

Two commonly used balancing strategies are stratified permuted block design and minimization, both of which require discrete covariates. In order to apply randomization procedures on continuous covariates in real clinical trials, the continuous

covariates must be discretized. A downside of discretizing continuous covariates is the lose of information, which can alter the nature of the covariates along with the influence the distribution balance.

Existing methods consider balancing continuous covariates based on a variety of aspects. Such aspects include (1) the differences of means and variances, (2) the weighted sum of the mean difference, standard deviation difference and the group size difference, (3) the area difference between empirical cumulative distribution functions, (4) the kernel density estimate distribution difference, and (5) the p-value of covariates across groups.

One way of handling continuous covariates is to use rank to replace actual covariate values. In this way continuous covariates can be treated in the same way as categorical covariates. The approach outlined in Ciolino (2011) measures the imbalance by ranking the pooled covariates and taking the ratio of the sum of ranks from the experimental treatment group to the sum of ranks from the control group. Another approach is to minimize the difference in rank-means, which is introduced in Hoehler (1987). Later Stigsby and Taves (2010) improved Hoehler's method by calculating rank-sum instead of rank-means. The rank-sum method first develops a rank-matrix constructed from an updated raw-matrix. This matrix represents the ranks of new subjects for individual prognostic variables. New subjects are tentatively assigned to either of the two available treatment groups. For each tentative assignment imbalance is calculated as the sum of squared deviations of the individual rank-sums for each prognostic variable.

Frane (1998) introduced a covariate-adaptive randomization for both continuous and categorical types that uses p-value to identify imbalance among treatment groups. A smaller p-value represents a larger imbalance among treatment groups. The p-value for continuous covariates is calculated using t-test and analysis of vari-

ance (ANOVA). For categorical covariates the goodness-of-fit χ^2 test is used. Let us consider a clinical trial with two groups (1 and 2) with the aim of balancing a single continuous prognostic variable. The first step is to temporarily assign a new patient into both groups and calculate the p-value for the t-test of the continuous covariates for each group. Subsequent patients are then randomized to group 1 with probability $p_1/(p_1 + p_2)$ and to group 2 with probability $p_2/(p_1 + p_2)$. The same process for randomized patient assignments can be generalized for n-prognostic variables and k-groups by letting q_{ij} denote the p-value for the j-th prognostic variable and the i-th group. The p-value for each group is determined by $p_i = min(q_{i1}, \cdots, q_{in})$. Overall Frane's method focuses on achieving balance between groups in relation to the prognostic variable with the greatest imbalance. However, a large p-values close to 1 for covariate values does not guarantee distributional balance between treatments. Since this method treats all covariates equally and imbalance is only based on the smallest p-value, it is not possible to rank covariates during clinical trials.

Endo (2006) intended to minimize the Kullback-Leibler divergence (KLD) as the index of difference in distribution between two groups, which is equivalent to a function of mean and standard deviation when the covariate is normally distributed. In a study, $f_{ij}(x)$ is the probability density function of X_{ij}, which indicates the j-th, $(j = 1, \cdots, J)$ prognostic variable of any subjects in group $i, (i = 1, 2)$. The difference in the distribution of prognostic variables between two groups can be expressed as an index of Kullback-Leibler divergence (KLD) using the equation below:

$$\delta_j = \int (f_{1j}(x) - f_{2j}(x))(logf_{1j}(x) - logf_{2j}(x))dx.$$

The smaller KLD indicates a smaller difference between two distributions, and thus can be interpreted to mean a higher degree of similarity. If the distribution is

assumed to be a normal distribution $N(\mu_{ij}, \sigma_{ij}^2)$, the KLD is:

$$\delta_j = 1/2\{(\mu_{1j} - \mu_{2j})^2 + (\sigma_{1j}^2 + \sigma_{2j}^2)\}(\frac{1}{\sigma_{1j}^2} + \frac{1}{\sigma_{2j}^2}) - 2.$$

If the prognostic variables are discrete random variables with probability function $p_{ij}(x)$, the KLD is:

$$\delta_j = \sum (p_{1j}(x) - p_{2j}(x))(logp_{1j}(x) - logp_{2j}(x)).$$

Finally, if the prognostic variable is a categorical variable of M_j categories, and the probability of each category are $\theta_{ij1}, \cdots, \theta_{ijM_j}$, the KLD is:

$$\delta_j = \sum_{m=1}^{M_j} (\theta_{1jm} - \theta_{2jm}) log \frac{\theta_{1jm}}{\theta_{2jm}}.$$

Before a new subject is allocated to each group, both $\delta_j^{(1)}$ and $\delta_j^{(2)}$ are calculated by temporarily assigning the new subject into two groups. The main limitation of this method is that, it is based on the probability density functions (PDFs) of the covariates. Compared to cumulative distribution functions it is more difficult to estimate PDFs. Probability density functions can be very close to zero, and KLD is calculated based on the ratio of the covariate's PDF in group 1 with the corresponding covariate's PDF in group 2. This may lead to an extreme outlier value of KLD, which can dominate the sum of the KLDs.

Su's quartile method (2011) does not have limitation on the type of covariates and it allows the covariates to be ranked according to their clinical importance, which can be perceived by the clinical trial practitioners. To determine the desired balance level for the group sizes and each of the selected covariates, let $C_0 = |S_t - S_c|$

denote the difference in the number of patients in two groups, where S_t denote the number of patients in treatment group and S_c denote the number of patients in control group. A binary imbalance score is defined as $C_0^b = I\{C_0 > c_0\}$, where $I\{\cdot\}$ is the indicator function and c_0 is a constant specified by the study team. For continuous symptom score, which is assumed to be positive, it is important to achieve similar overall distributions in both groups by maintaining similar quartiles for both groups. By denoting the three quartiles for the treatment and control groups as $Q_i^t, i = 1, 2, 3$ and $Q_i^c, i = 1, 2, 3$, respectively, the continuous imbalance score can be defined as

$$C_1 = max\{|Q_i^t - Q_i^c|/max(Q_i^t, Q_i^c), i = 1, 2, 3\}.$$

The binary score is then

$$C_1^b = I\{max\{|Q_i^t - Q_i^c|/max(Q_i^t, Q_i^c), i = 1, 2, 3\} > c_1\},$$

where c_1 is the threshold for imbalance. For categorical prognostic variable, the imbalance score can be defined as $C_2 = max|N_i^t - N_i^c|, i = 1, 2, \cdots$, where N_i^t and N_i^c denote the number of patients in covariate i for the treatment and control groups, respectively. Similarly, the imbalance score can be denoted as $C_2^b = I\{\max\{|N_i^t - N_i^c|, i = 1, 2, \cdots\} > c_2\}$. The importance of the factors is ordered, and a weighted overall imbalance score is based on the following binary imbalance score:

$$C = w_0 C_0^b + w_1 C_1^b + w_2 C_2^b,$$

where $w_i, i = 0, 1, 2$ are the weights assigned to the factors with $w_0 > w_1 > w_2$. The new patient should be assigned to the treatment that will have the smaller

continuous imbalance score. Su's method uses the largest difference in the three pairs of quartiles of the two distributions of the observed covariate values to quantify the imbalance level of continuous covariate. However, the two distributions can also be significantly different even with the same quartiles.

Motivated by attaining identical distribution functions at the end of a trial, Lin and Su (2012) use the normalized area between two empirical cumulative distribution functions (ECDFs) as the imbalance measure of the distribution of the covariates. For continuous covariates, Lin and Su denote the ECDFs for the treatment and control group by $\hat{F}(t)$ and $\hat{G}(t)$, respectively, where

$$\hat{F}(t) = \frac{1}{m}\sum_{i=1}^{m} I\{x_i \leq t\}, \hat{G}(t) = \frac{1}{n}\sum_{i=1}^{n} I\{y_i \leq t\},$$

and define the area between $\hat{F}(t)$ and $\hat{G}(t)$ as $A(\hat{F}(t), \hat{G}(t))$. The normalized imbalance metric is:

$$\tilde{A}(\hat{F}(t), \hat{G}(t)) = \frac{A(\hat{F}(t), \hat{G}(t))}{max(x_1, \cdots, x_m, y_1, \cdots, y_n) - min(x_1, \cdots, x_m, y_1, \cdots, y_n)},$$

and is bounded between 0 and 1, representing the perfect balance and the worst imbalance, respectively.

For a categorical covariate with $K \geq 2$ different categories or levels, there is an experimental and a control treatment group. The percent of m patients in the K categories receiving the experimental treatment is $p_1, \cdots, p_K$. In the control treatment, the percent of n patients in K categories is $q_1, \cdots, q_K$. The ideal balance is achieved when $p_1 = q_1, \cdots, p_K = q_K$. The area between ECDF of covariate B_i is

$|p_i - q_i|$. The sum of the areas for the K categories is

$$\sum_{i=1}^{K} |p_i - q_i| \leq \sum_{i=1}^{K} (p_i + q_i) = 2$$

with equality holding, if and only if, at least one of p_i and q_i equal to 0 for any i. For a categorical covariate, the normalized area is

$$\tilde{A} = \sum_{i=1}^{K} |p_i - q_i|/2.$$

Ma and Hu (2013) proposed to balance continuous covariates by minimizing the differences between covariate densities, which can be estimated using a kernel density estimator. They consider a continuous covariate Z, and let $Z_{i,k}, i = 1, ..., n_k$, denote the baseline scores of patients in treatment $k, k = 1, 2$. The density function for covariate Z in treatment k can be estimated using the kernel method as follows:

$$\hat{f}_k(z) = \frac{1}{n_k h(n_k)} \sum_{i=1}^{n_k} K(\frac{z - Z_{i,k}}{h(n_k)}),$$

where $K(\cdot)$ is the kernel function. Ma and Hu use Scott's rule: $h(n_k) = \hat{\sigma} n_k^{-0.2}$, where $\hat{\sigma}$ is the estimated standard deviation of the covariate. Distributional imbalance covariates is defined below:

$$\triangle d = \frac{n_1}{n} \hat{f}_1(z_0) - \frac{n_2}{n} \hat{f}_2(z_0),$$

where z_0 is the baseline score of a new patient. More generally, $\triangle d_j$ denotes the distributional imbalance of covariate $j, j = 1, \cdots, M$. The overall imbalance measure

Imb is defined as the weighted average of M covariate imbalances:

$$Imb = \sum_{j=1}^{M} w_j \triangle d_j,$$

where w_j is a nonnegative weight on covariate j. And the allocation rule is: assign new patient to treatment 1 with probability π, if $Imb < 0$; assign to treatment 2 with probability π, if $Imb > 0$; and assign with probability 1/2, if $Imb = 0$. For this method, the maximum acceptable imbalance of group size does not need to be pre-specified. All other adaptive designs for continuous covariates need the threshold imbalance for $|n_1 - n_2|$, and when it is reached the patient is allocated without randomization.

1.3 Testing Hypotheses of covariate-adaptive design

In practice, conventional testing hypotheses are employed without consideration of covariate-adaptive randomization scheme. The validity of conventional tests under covariate-adaptive designs is still a concern. Forsythe (1987) suggests that all covariates used in minimization procedure should also be incorporated in inference models. Shao, Yu and Zhong (2010) found one way to obtain a valid test procedure is to use a correct model between outcomes and covariates, including those used in randomizations. It is well known now that covariates used in trial randomization should also be incorporated in inference procedures. A valid test procedure should be associated with the randomization scheme. Although covariate-adaptive design has long been in use, hypothesis testing under these designs is rarely discussed. Most works that do discuss hypothesis testing are related to discrete covariates.

Meanwhile, hypothesis testing for continuous covariates are even less understood. Thus it is very important to develop a testing that is valid under covariate-adaptive design that balances both discrete and continuous covariates. The test should also be more powerful under covariate-adaptive randomization than it is under simple randomization. Discussion based on simulations can be found in the literature (see, Brikett, 1985; Forsythe, 1987; Aickin, 2002; Weir and Lees, 2003; Hagino, 2004 and Shao et al., 2010). Ma and Hu (2014) establish a theoretical framework for hypotheses testing of adaptive designs with discrete covariates under linear model. In this paper, we propose a theoretical foundation about hypotheses testing regards under adaptive designs with continuous covariates.

1.4 Motivation of my dissertation

Clinical trials are always complex with multiple objectives. Common objectives include maximizing power to detect clinically relevant difference, balancing important covariates for valid comparison, detecting important interaction among treatments and covariates, and minimizing total monetary cost of a trial. Covariate-adaptive randomization methods achieve a balanced treatment allocation over important covariates. In this section we will discuss the applicability and limits of existing methods with regards to handling continuous covariates in clinical trials. We also briefly introduce our work in studying and understanding adaptive designs and inference with continuous covariates under conventional hypothesis testing.

The two most commonly used balancing strategies are stratified permuted block design and minimization, both of which require the covariates to be discrete. In practice, there are many continuous covariates which need to be discretized in order to be included in randomization procedures. However, lose of information is a

consequence of breaking down continuous covariate into subcategories, which may change the nature of the covariates and influence the distribution balance.

Existing adaptive designs for continuous covariates each have their own strengths and weakness. Most adaptive randomizations can deal with certain individual aspects of continuous covariates. Meanwhile only Ma and Hu (2013) have targeted the density using the kernel estimator. Hoehler (1986) and Stigsby focus on rank, Frane (1988), Nishi (2003) and Endo (2006) focus on mean and standard deviation, Su (2011) focus on three quartiles, and Lin and Su (2012) focus on empirical cumulative distribution. Although each of criterions they use can represent the distribution in some aspect, none of them can ensure the same distribution. For example, Endo (2006) shows that similarity in mean and standard deviation ensures similar distributions when covariates are normally distributed. Realistically, baseline factors are not normally distributed and therefore the normality of covariates in each group cannot be properly justified. In fact, Endo's method can possibly split the normal distribution to a left-skewed sample and a right-skewed sample with the same mean and standard deviation. Ma and Hu's kernel estimate method comprehensively balances the distribution of the covariates without having to worry about normality. Even though Ma and Hu's targets the distribution of covariates, it is not a minimization procedure. Therefore this work modifies the Ma and Hu's method to be a minimization procedure.

Our proposed version of the kernel density continuous covariate randomization method follows the idea of minimization and out-performs existing covariate-adaptive designs in balancing the distribution difference. This approach also minimizes the difference of mean, median, variance and other characteristics as those continuous covariate-adaptive designs targeting on that specific characteristics.

The state-of-the-art work on performance strategies that balance continuous co-

variates are all based on simulation. None of the current works provide a theoretical understanding of the asymptotical properties of adaptive designs for continuous covariates. Thus conventional hypothesis testing for clinical trials using continuous covariate-adaptive designs is not well understood. Herein, this dissertation investigates the theoretical properties of the testing hypotheses by proving the asymptotic distributions of the test statistic under null and alternative hypotheses for continuous covariate-adaptive designs that satisfy the conditions presented in theorem 2.2.1. In Chapter 2 and 3, we also did extensive simulation studies on a class of continuous covariate-adaptive randomization methods, including the rank-sum test, the p-value based method, the Su's percentile method, the empirical cumulative-distribution method, the Kullback-Leibler divergence method, and the kernel-density method. The Type-I error and power of the hypothesis testing under those adaptive designs are compared to those using the complete randomization method.

Chapter 2

Statistical inference for adaptive designs balance independent continuous covariates

2.1 Background and notations

A statistical hypothesis test is a method of statistical inference used for testing a statistical hypothesis. Given a threshold probability---the significance level, a test result is statistically significant if it has been predicted as unlikely to happen by sampling error alone. In the Neyman-Pearson framework, the process of distinguishing the null hypothesis from the alternative hypothesis is by identifying type I and type II error. It is essential to make sure the hypothesis testing is valid when comparing two treatment effects. A hypothesis testing is valid under the following condition: $lim_{N\to\infty} P_{y,I}\left(|T| > C_\alpha | \mathcal{Z}\right) \leq \alpha$ with equality holding for at least some cases. Meanwhile a test is conservative when the true probability of incorrectly rejecting the null hypothesis is never greater than the nominal level when constructed for a given

nominal significance level. The result of hypothesis testing is not convincing if the test is conservative.

The hypothesis test here is based on linear model framework under covariate-adaptive designs. Assuming there is a clinical trial with two treatments: 1 and 2, the discrete and continuous covariates of patients are balanced by applying one covariate-adaptive design. Let μ_j, $j = 1, 2$ is the expected main effect of treatment j respectively; N is the total number of patients in this trial; I_i $i = 1, \cdots, N$ is the allocation indicator of patient i, say $I_i = 1$ if patient i is assigned to treatment 1, $I_i = 0$ otherwise. Assume the relationship between the response Y and covariates, treatment follows the linear model below:

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + \alpha_1 X_{i,1} + \cdots + \alpha_p X_{i,p} + \beta_1 Z_{i,1} + \cdots + \beta_q Z_{i,q} + \varepsilon_i,$$

where

1. $X_{i,k}$, $k = 1, \cdots, p$ is discrete or continuous covariate identically independent distributed as X_k, with $E(X_k) = 0$, which is used in both covariate randomization procedure and final statistical inference.

2. $Z_{i,j}$, $j = 1, \cdots, q$ is discrete or continuous covariate identically independent distributed as Z_j, with $E(Z_j) = 0$, which is only used in covariate randomization procedure.

3. ε_i, $i = 1, \cdots, N$ is independent and identically distributed random error with $E(\varepsilon_i) = \sigma_\varepsilon^2$.

4. $X_{i,k}$, $k = 1, \cdots, p$ and $Z_{i,j}$, $j = 1, \cdots, q$ are independent from each other, and ε_i is independent with $X_{i,k}$ and $Z_{i,j}$.

Let us define $\tilde{Y} = (Y_1, Y_2, \cdots, Y_N)^T$, $\tilde{\alpha} = (\mu_1, \mu_2, \alpha_1, \cdots, \alpha_p)^T$, $\tilde{\beta} = (\beta_1, \beta_2, \cdots, \beta_q)^T$ and $\tilde{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N)^T$,

$$\mathbf{X} = \begin{bmatrix} I_1 & (1-I_1) & X_{1,1} & \cdots & X_{1,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I_N & (1-I_N) & X_{N,1} & \cdots & X_{N,p} \end{bmatrix}, \ \mathbf{Z} = \begin{bmatrix} Z_{1,1} & \cdots & Z_{1,q} \\ \vdots & \ddots & \vdots \\ Z_{N,1} & \cdots & Z_{N,q}. \end{bmatrix}.$$

Notice both $X_{i,k},\ k = 1, \cdots, p$ and $Z_{i,j},\ j = 1, \cdots, q$ are assumed to be scalar quantities. For covariates with two categories, a dummy variable is used. In cases with more than two categories, for example when there are three categories, high dimensional vectors (0,0), (0,1), and (1,0) are coded in the model. All the results can be extended to the situation that discrete covariates have multiple categories easily.

The general model for randomization can be rewritten as,

$$\tilde{Y} = \mathbf{X}\tilde{\alpha} + \mathbf{Z}\tilde{\beta} + \tilde{\varepsilon}. \tag{2.1}$$

The statistical inference working model is then,

$$\tilde{Y} = \mathbf{X}\tilde{\alpha} + \tilde{\varepsilon}. \tag{2.2}$$

Hypothesis testing: Based on the final statistical inference model 2.2, the following hypothesis test is used to compare if there is difference between two treatment

effects:

$$H_o : \mu_1 = \mu_2 \; vs. \; H_a : \mu_1 \neq \mu_2. \tag{2.3}$$

According to the ordinary least square estimation method, the estimate of $\tilde{\alpha}$ is obtained by the following formula,

$$\hat{\alpha} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\tilde{Y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\tilde{\alpha} + \mathbf{Z}\tilde{\beta}).$$

The test statistic for (2.3) is

$$T = \frac{L\hat{\alpha}}{(\hat{\sigma}^2 L(\mathbf{X}^T\mathbf{X})^{-1}L^T)^{1/2}}, \tag{2.4}$$

where $L = (1, -1, 0, \cdots, 0)$ and $\hat{\sigma}^2 = (\tilde{Y} - \mathbf{X}\hat{\alpha})/(N - p - 2)$. The null hypothesis is rejected when $|T| > Z_{1-\alpha/2}$ and otherwise accepted, where $Z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ percentile of the standard normal distribution.

Evaluating the significance of a single covariate is usually important in personalized medicine. To test the significance of a single covariate, without loss of generality, we consider the hypothesis testing for α_1 using the hypothesis,

$$H_0 : \alpha_1 = 0 \; vs. \; H_a : \alpha_1 \neq 0. \tag{2.5}$$

The test statistic for hypothesis testing (2.5) is:

$$T = \frac{(l\hat{\alpha})}{(\hat{\sigma}^2 l(\mathbf{X}^T\mathbf{X})^{-1}l^T)^{1/2}}, \tag{2.6}$$

where $l = (0, 0, 1, 0, \ldots, 0)$. The null hypothesis is rejected if $|T| > Z_{1-\alpha/2}$, otherwise

the null hypothesis is accepted.

In general forms of hypothesis testing for significance of covariates, let C be an $m \times (p+2)$ matrix of rank m with $m < (p+2)$, where entries of the first two columns are all zeros. The hypothesis test would be

$$H_0 : C\tilde{\alpha} = \xi_0 \; vs. \; H_a : C\tilde{\alpha} = \xi_1. \tag{2.7}$$

The test statistic for hypothesis testing (2.7) is,

$$T^* = \frac{m^{-1}(C\hat{\alpha} - \xi_0)^T [C(\mathbf{X}^T\mathbf{X})^{-1}C^T]^{-1}(C\hat{\alpha} - \xi_0)}{\hat{\sigma}^2}. \tag{2.8}$$

The null hypothesis is rejected when $T^* > \chi^2_{m,(1-\alpha)}/m$, where $\chi^2_{m,(1-\alpha)}/m$ is $(1-\alpha)$ percentile of a χ^2 distribution with degree of freedom m, otherwise, accept the null hypothesis.

2.2 Asymptotic properties

Suppose a clinical trial is designed to balance both discrete and continuous covariates: $X_{i,k},\ k = 1, \cdots, p,\ Z_{i,j},\ j = 1, \cdots, q$. Let $D = (k|X_k$ is categorical, $k = 1, ..., p)$, $C = (k|X_k$ is continuous, $k = 1, ..., p)$, $D^* = (j|Z_j$ is categorical, $j = 1, ..., q)$ and $C^* = (j|Z_j$ is continuous, $j = 1, ..., q)$. The marginal imbalance measure for all levels of categorical covariates are considered here. Assume categorical covariate $X_k \in D$ has level s_k and $Z_j \in D^*$ has level s_j^*. Continuous covariate $X_k \in C$, $Z_k \in C^*$ whose two group difference is define as the difference of the sum of this covariate in two groups. For i^{th} patient, we use $W_i = (X_{i,1}, X_{i,2}, ..., X_{i,p}, Z_{i,1}, Z_{i,2}, ..., Z_{i,q})$ to represent the covariate profile. If $X_{i,k}$ is categorical and at level t_k $(1 \leq t_k \leq s_k)$ and $Z_{i,j}$ is categorical and at level t_j

$(1 \leq t_j \leq s_j^*)$, for convenience, we use $(k; t_k)$ to denote the margin at categorical level $X_k = t_k$, and $(j; t_j)$ to denote the margin at categorical level $Z_j = t_j$. The overall and marginal imbalance between two treatments are defined as:

1. D_N be the difference between the number of patients among two groups as total, where $D_N = N_1 - N_2$.

2. $D_N(k; t_k)$, $D_N(j; t_j)$ be the differences between the number of patients in the two treatment groups on the margin $(k; t_k)$, $(j; t_j)$ for categorical covariates, where $D_N(k; t_k) = N_{t_k,1} - N_{t_k,2}$, $D_N(j; t_j) = N_{t_j,1} - N_{t_j,2}$.

3. $D_N^c(X_k)$, $D_N^c(Z_j)$be the difference between continuous covariate k and j among two groups, where $D_N^c(X_k) = \sum_{i_1=1}^{N_1} X_{i_1,k} - \sum_{i_2=1}^{N_2} X_{i_2,k}$, $D_N^c(Z_j) = \sum_{i_1=1}^{N_1} Z_{i_1,j} - \sum_{i_2=1}^{N_2} Z_{i_2,j}$.

The primary interest of hypothesis testing is comparing two treatment group effect difference, and the secondary interest is testing significance of covariates, both types of hypothesis testing are considered here. For clinical inference the working model is (2.2) while the data are generated from the true model (2.1). Asymptotic properties of testing statistic under both null hypothesis and alternative hypothesis is studied here.

Theorem 2.2.1. *Suppose the following three conditions are satisfied in a covariate-adaptive design:*
(1) the overall imbalance converges to zero in probability at rate $N^{1/2}$, that is $D_N = o_p(N^{1/2})$;
(2) the marginal imbalance for each categorical covariate converges to zero in probability at rate $N^{1/2}$, that is, $D_N(k, t_k) = o_p(N^{1/2})$ and $D_N(j, t_j) = o_p(N^{1/2})$;
(3) each continuous covariate sum in two groups converges to zero in probability at

rate $N^{1/2}$*, that is,* $D_N^c(X_k) = o_p(N^{1/2})$ *and* $D_N^c(Z_j) = o_p(N^{1/2})$*.*

(i) Then under $H_0 : \mu_1 - \mu_2 = 0$*,*

$$T \xrightarrow{D} N(0, \tau^2) \ , where \ \ \tau^2 = \frac{\sigma_\varepsilon^2}{\sigma_z^2}, \ \sigma_z^2 = \sigma_\varepsilon^2 + \sum_{j=1}^{q} \beta_j^2 Var(Z_j), \tag{2.9}$$

if $\beta_j = 0$ *for all* $j = 1, ..., q$*, then* $\tau = 1$*. Thus only if all the covariates used in randomization are not related with outcome* Y*, the hypothesis testing can achieve type I error, otherwise, the hypothesis testing is conservative.*
(ii) Under $H_a : \mu_1 - \mu_2 \neq 0$*, consider a sequence of local alternative, i.e,* $\mu_1 - \mu_2 = \delta/\sqrt{N}$ *for a fixed* $\delta \neq 0$*, then*

$$T \xrightarrow{D} N(\Delta, \tau^2), \quad where \ \ \Delta = \frac{\delta}{2\sigma_z}. \tag{2.10}$$

Theorem 2.2.1, gives the theoretical properties of test statistic for testing hypothesis of treatment effects under covariate-adaptive designs. Covariate information is used in covariate-adaptive randomization to reduce the imbalance of different levels of discrete covariates (overall, marginal, and within-stratum) and certain characteristics of continuous covariates (mean, variance, quantiles, distribution). Under the assumption of three mild conditions of covariate-adaptive designs, the asymptotic distribution of the test statistic is derived. If the overall two group difference converges to zero at rate $N^{1/2}$, the marginal difference of categorical covariates at all levels, and the continuous covariate sum goes to zero at rate $N^{1/2}$. These conditions are satisfied by most well-known covariate-adaptive designs. The proof shows under these three conditions, the model based variance used in the denominator of test statistic is smaller than its actual variance. If the important covariates are omit-

ted from the working model, type I error is smaller than nominal level under null hypothesis.

Under the alternative hypothesis, the power can be obtained from the asymptotic distribution 2.10. Under covariate-adaptive design, power is

$$P(|T| > Z_{1-\alpha/2}) = \Phi(\frac{\delta}{2\sigma_\varepsilon} - \frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\varepsilon}) + \Phi(-\frac{\delta}{2\sigma_\varepsilon} - \frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\varepsilon}) + o(1).$$

The power of complete randomization would be

$$P(|T| > Z_{1-\alpha/2}) = \Phi(\frac{\delta}{2\sigma_\varepsilon} - Z_{1-\alpha/2}) + \Phi(-\frac{\delta}{2\sigma_\varepsilon} - Z_{1-\alpha/2}) + o(1),$$

because $\sigma_\varepsilon = \sigma_z$ under complete randomization. Some conclusions about the power comparisons of covariate-adaptive designs and complete randomization can be made. The asymptotic power under covariate-adaptive design should be smaller than complete randomization when δ is small, and usually larger than complete randomization when δ is large. The simulation results also confirm these conclusions.

The hypothesis testing of the significance of covariates can still be valid in perspective of type I error, even though not all covariates used in randomization are incorporated in the inference model. The power will also be harmed if important covariates which used in randomization procedure is missed in the inference model.

Theorem 2.2.2. *Under the same three conditions as in Theorem 2.2.1*
(i) under null hypothesis $H_0 : C\tilde{\alpha} = \xi_0$,

$$T^* \xrightarrow{D} \chi^2_{(m)}/m \tag{2.11}$$

where $T^* = m^{-1}(C\hat{\alpha} - \xi_0)^T[C(\mathbf{X}^T\mathbf{X})^{-1}C^T]^{-1}(C\hat{\alpha} - \xi_0)/\hat{\sigma}^2$.

(ii) Under $H_a : C\tilde{\alpha} = \xi_1$, *if* $(\xi_1 - \xi_0) = \eta/\sqrt{N}$

$$T^* \xrightarrow{D} \chi^2_{(m)}(\lambda)/m,\ \lambda = \eta^T[CM^{-1}C^T]^{-1}\eta/\hat{\sigma}^2_Z. \tag{2.12}$$

where $M = diag(1/2, 1/2, \mathsf{Var}(X_1), ..., \mathsf{Var}(X_p))$ *and* λ *is the noncentral parameter.*

Theorem 2.2.2 shows that the hypothesis tests regarding significance of covariates can still obtain the correct type I error, under covariate-adaptive design. This suggests that to test the significance of some covariate in real clinical trials, the inference model can only contain partial covariates. Under covariate-adaptive design, however, the power of hypothesis testing of covariates will decrease by omitting any important covariate in the inference model.

The following corollary gives an important special case of hypothesis testing of covariates, where only one covariate is considered.

Corollary 2.2.1. *Under the same three conditions as in Theorem 2.2.1,*
under $H_0 : \alpha_1 = 0$,

$$T_1 \xrightarrow{D} N(0,1),\ where\ T_1 = \frac{l\hat{\alpha}}{(\hat{\sigma}^2 l(\mathbf{X}^T\mathbf{X})^{-1}l^T)^{1/2}} \tag{2.13}$$

under $H_a : \alpha_1 \neq 0$, *i.e.* $\alpha_1 = \delta_{\alpha_1}/\sqrt{N}$.

$$T_1 \xrightarrow{D} N(\Delta_{\alpha_1}, 1),\ where\ \Delta_{\alpha_1} = \frac{\delta_{\alpha_1}\sigma_{X_1}}{\sigma_z}. \tag{2.14}$$

Corollary 2.2.1 shows the hypothesis testing of a single covariate can still achieve correct type I error with incomplete inference model. And the power will be harmed if missing important covariates and will increase as more covariates are incorporated into the model.

2.3 One example: adaptive design balances mean difference

Assume there is a clinical trial with two treatments, 1 and 2. Mean difference based method (Mean-diff) is used to balance one continuous covariate Z. Sample $Z_1, Z_2, \cdots, Z_n$ are independent and identically distributed with mean 0 and variance σ^2. Suppose n_1, n_2 patients are assigned to treatment 1, 2, respectively. $Z_{1,1}, Z_{2,1}, \cdots, Z_{n_1,1}$ are assigned to group 1, and $Z_{1,2}, Z_{2,2}, \cdots, Z_{n_2,2}$ are assigned to group 2. The imbalance measure is defined as the square of mean difference:

$$Imb_n = (\overline{Z}_{n_1} - \overline{Z}_{n_2})^2.$$

When the $(n+1)$-th patient enters the trial with covariate Z_{n+1}, we determine the allocation of this new patient using a biased coin, which favors the treatment with smaller mean difference between two groups. Let p be a biased probability, such that $0.5 < p < 1$, the procedure of biased coin allocation for continuous covariate can be summarized as follows.

Step 1 : Initial step

Assign n_0 patients to each treatment by using a restricted randomization.

Step 2 : Imbalance calculation

Suppose the $(n+1)$-th patient is potentially assigned to treatment k, $k = 1, 2$; calculate the imbalance measure $Imb^{(1)}$ and $Imb^{(2)}$.

Step 3 : Biased allocation

(a) $|n_1 - n_2| < d$

If $Imb^{(1)} < Imb^{(2)}$, assign patient $n+1$ to treatment 1 with probability

p.

If $Imb^{(1)} > Imb^{(2)}$, assign patient $n+1$ to treatment 2 with probability p.

If $Imb^{(1)} = Imb^{(2)}$, assign patient $n+1$ to treatment 1 or 2 with probability 1/2.

(b) $|n_1 - n_2| = d$

assign $(n+1)$-th patient to treatment with less patients with probability 1.

Theorem 2.3.1. *Under adaptive design based on mean difference,* $\{\Lambda_n = \sum_{i=1}^{N_1} Z_{i1} - \sum_{i=1}^{N_2} Z_{i2}\}$ *is a positive recurrent Markov chain, which means* $\Lambda_n = O_p(1)$. *Thus this adaptive randomization is one example that satisfies three conditions in Theorem 2.2.1.*

2.4 Simulation study

Case 1: Testing treatment effects. Continuous covariate randomization methods: mean difference based method (Mean-diff), p-value based randomization (P-value) by Frane (1998), Su's (2011) percentile method (Quartile), Empirical cumulative distribution function method (ECDF) by Lin and Su (2012), Kullback-Leibler divergence method (KLD) by Endo et al. (2006) and Kernel density procedure (Kernel) by Ma and Hu (2013) are compared with complete-randomization (CR) to compare the type I error of the hypothesis testing for comparing treatment effects. The details of these adaptive designs are in chapter 1.2.2.

To investigate the type I error of the hypothesis testing: $H_o: \mu_1 = \mu_2$, no treatment effect difference is assumed here, i.e., $\mu_1 = \mu_2$. In the simulation, biased coin

probability $p = 0.8$ is applied for all the covariate-adaptive designs. The significance level $\alpha = 0.05$ and sample size $N = 100, 200, 500$ is considered. The groups size difference is set to be no more than 6. Three types of tests are compared: (1) two sample t-test (t-test), (2) covariate test based on the linear model contains covariate Z ($lm(z), lm(z_1, z_2)$), (3) bootstrap t-test introduced in Shao, Yu, and Zhong (2010), where bootstrap samples $(Y_1^{*b}, Z_1^{*b}), \cdots, (Y_N^{*b}, Z_N^{*b})$, $b = 1, 2, \cdots, B$, are generated independently randomly with replacement from sample $(Y_1, Z_1), \cdots, (Y_N, Z_N)$. The variance of $\bar{Y}_1 - \bar{Y}_2$ is estimated by the bootstrap samples.

Model 1: the response Y_i is assumed to follow linear model:

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 Z_i + \varepsilon_i, \tag{2.15}$$

where $Z_i \sim N(0, 1)$ and $\varepsilon_i \sim N(0, 1)$, $\beta_1 = 1$.

Table 2.1: Type I error for methods: mean difference based method (Mean-diff), p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR) in %, simulation based on 10,000 runs, significance level $\alpha = 5\%$ and sample size $N = 100, 200, 500$.

Method	N	t-test	$lm(z)$	B-test
Mean-diff	100	0.66	4.92	5.34
	200	0.61	5.15	5.12
	500	0.62	4.95	4.69
P-value	100	0.01	5.12	5.16
	200	0.02	4.82	4.45
	500	0.01	5.13	5.26
Quartile	100	1.38	5.12	5.85
	200	1.26	5.05	4.68
	500	1.00	4.85	5.53
ECDF	100	0.75	4.84	4.56
	200	0.93	5.16	5.27
	500	0.63	4.89	4.68
KLD	100	0.02	5.02	4.85
	200	0.01	5.56	5.86
	500	0.00	4.86	4.73
Kernel	100	1.09	5.31	5.09
	200	0.80	4.76	5.69
	500	0.67	5.00	4.53
CR	100	4.63	4.75	-
	200	5.16	5.05	-
	500	4.59	4.81	-

Model 2 considers about a linear model with two covariates Z_1, Z_2, the response Y_i is assumed follows linear model:

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + \varepsilon_i \tag{2.16}$$

where $Z_{i,1} \sim N(0,1)$, $Z_{i,2} \sim exp(1)$ and $\varepsilon_i \sim N(0,1)$, $\beta_1 = \beta_2 = 1,\ \mu_1 = \mu_2 = 100$.

Table 2.2: Type I error for methods: p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR) in %, simulation based on 10,000 runs, significance level $\alpha = 5\%$ and sample size $N = 100, 200, 500$.

Method	N	t-test	$lm(Z_1)$	$lm(Z_2)$	$lm(Z_1, Z_2)$	B-test
P-value	100	0.13	0.82	0.71	4.98	4.49
P-value	200	0.08	0.61	0.59	4.96	4.73
P-value	500	0.05	0.59	0.62	5.16	5.67
Quartile	100	1.36	1.85	2.43	4.99	5.03
Quartile	200	1.07	1.43	2.39	4.96	5.09
Quartile	500	1.44	1.31	2.57	4.79	5.53
ECDF	100	0.28	1.21	0.88	5.09	4.67
ECDF	200	0.19	0.79	0.76	4.61	5.49
ECDF	500	0.12	0.67	0.61	4.54	5.20
KLD	100	0.32	1.24	0.95	5.51	4.93
KLD	200	0.14	0.79	0.68	5.35	5.45
KLD	500	0.06	0.64	0.65	5.12	5.63
Kernel	100	0.76	1.73	1.41	5.06	4.85
Kernel	200	0.55	1.33	1.23	5.06	4.89
Kernel	500	0.35	1.39	1.01	5.25	5.57
CR	100	4.54	4.70	5.11	5.14	-
CR	200	4.86	4.93	4.84	5.61	-
CR	500	5.13	4.76	5.28	5.26	-

Based on Table 2.1, several conclusions can be drawn. Firstly, the two sample t-test is conservative under all the methods here except complete randomization. This also means t-test is only valid under complete randomization, and conservative for restricted randomization procedures. Secondly, under full model $lm(Z)$ is valid for all the randomization procedures, thus it is more powerful than t-test. From Table 2.2, it is obvious that t-test, $lm(Z_1)$ and $lm(Z_2)$ are all conservative. Among them, the two sample t-test is the most conservative with smallest type I error. $lm(Z_1)$ or $lm(Z_2)$ is more conservative than $lm(Z_1, Z_2)$, which indicates that missing important covariates will damage the validness of the test. In addition, the bootstrap method is always valid for covariate-adaptive designs in both cases.

Case 2: Power comparison. To the comparison of power for different hypothesis testing methods under continuous covariate-adaptive designs and complete randomization, same linear model with one covariate is used as in case 1, but the difference between treatment effects μ_1 and μ_2 is not zero. The response Y_i is assumed to follow linear model:

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 Z_i + \varepsilon_i, \tag{2.17}$$

where $Z_i \sim N(0,1)$ and $\varepsilon_i \sim N(0,1)$, $\beta_1 = 1, \mu_1 - \mu_2 = \frac{i}{10}, i = 1, 2, \cdots, 10$. We set $\mu_1 - \mu_2$ =0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0, sample size $N = 32, 64, 100$. The simulation repeated 10,000 times, each time biased coin probability $p = 0.8$ is used to minimize the imbalance measure.

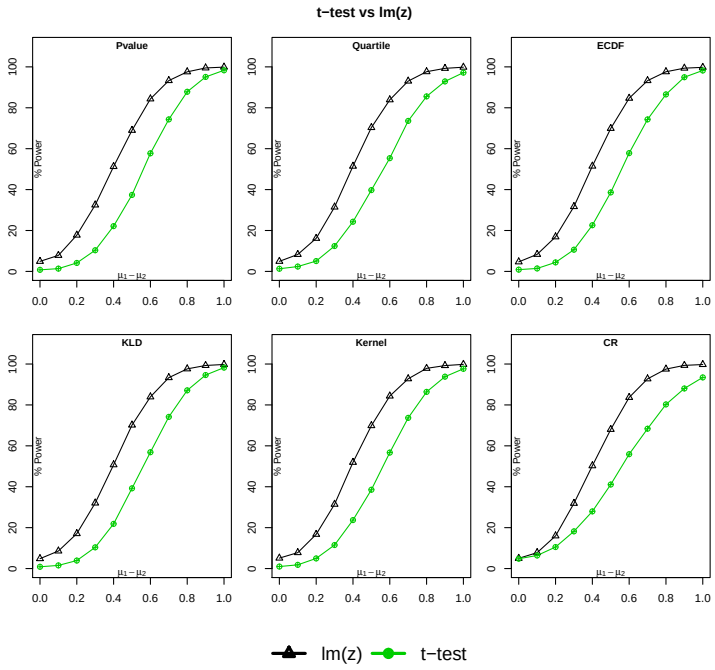
Table 2.3: Power Comparison for methods: mean difference based method (Mean-diff), p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR) in %, simulation based on 10,000 runs and sample size is 32.

Test	$\mu_1 - \mu_2$	Mean-diff	P-value	Quartile	ECDF	KLD	Kernel	CR
t-test	0	0.86	0.87	1.41	1.16	1.21	1.58	4.58
	0.1	1.07	1.24	1.76	1.45	1.21	1.79	5.68
	0.2	1.89	1.99	2.54	2.19	2.14	2.71	6.96
	0.3	3.25	3.31	4.06	3.59	3.34	4.49	9.38
	0.4	5.45	5.90	6.81	6.02	6.00	7.07	11.51
	0.5	8.68	8.92	10.34	9.43	9.58	10.77	15.26
	0.6	13.90	13.69	15.99	15.79	14.59	14.92	20.97
	0.7	20.69	20.36	21.92	20.91	20.90	21.70	26.54
	0.8	27.76	27.79	29.58	28.36	28.90	29.33	31.86
	0.9	37.33	36.21	37.62	37.29	38.46	38.38	39.85
	1	47.28	46.94	47.54	47.34	47.08	47.73	47.15
$lm(z)$	0	5.15	4.98	5.13	5.33	5.67	5.09	5.18
	0.1	5.89	6.18	5.89	6.23	5.33	6.18	5.81
	0.2	8.25	8.49	8.19	8.68	8.77	8.40	8.28
	0.3	12.61	12.96	12.92	12.89	12.46	12.80	12.69
	0.4	19.06	19.28	19.86	18.92	19.63	19.26	18.21
	0.5	26.53	27.89	26.91	27.12	27.38	27.24	25.07
	0.6	36.93	36.17	37.81	37.80	36.29	37.12	34.57
	0.7	47.32	48.01	47.43	47.45	48.05	47.14	45.16
	0.8	58.07	57.01	58.57	58.11	58.15	57.21	54.58
	0.9	69.07	67.34	68.07	68.59	69.12	68.69	65.12
	1	77.81	77.26	76.89	77.65	76.85	77.75	74.43

Table 2.4: Power Comparison for methods: mean difference based method (Mean-diff), p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR) in %, simulation based on 10,000 runs and sample size is 64.

Test	$\mu_1 - \mu_2$	Mean-diff	P-value	Quartile	ECDF	KLD	Kernel	CR
t-test	0	0.76	0.67	1.33	0.95	0.74	1.30	5.62
	0.1	1.15	1.12	2.06	1.45	1.34	1.60	6.20
	0.2	2.62	2.58	3.95	2.97	3.06	3.91	8.85
	0.3	5.83	6.11	8.28	6.43	6.40	7.27	12.51
	0.4	12.27	11.83	14.15	13.10	12.98	13.42	19.61
	0.5	21.64	21.13	22.95	22.36	21.94	23.06	27.74
	0.6	33.19	34.05	35.43	34.57	34.13	35.30	37.99
	0.7	48.86	49.01	49.17	48.89	49.02	48.91	48.82
	0.8	63.57	64.05	62.80	63.81	63.77	62.75	60.36
	0.9	77.46	77.12	74.64	76.71	76.69	75.17	69.47
	1	86.32	86.72	85.05	86.21	86.46	85.07	78.95
$lm(z)$	0	5.44	5.03	4.99	5.04	4.90	4.84	5.27
	0.1	6.57	6.56	7.21	6.98	7.02	6.98	6.72
	0.2	12.16	12.03	12.43	12.26	12.37	12.39	12.48
	0.3	21.39	21.74	22.18	21.34	21.68	21.38	21.08
	0.4	35.60	34.48	35.05	34.98	35.43	34.89	34.45
	0.5	50.65	49.74	49.55	50.20	50.01	50.03	48.52
	0.6	65.04	66.38	64.79	64.85	66.37	66.00	64.89
	0.7	78.93	78.50	78.12	78.75	78.91	78.56	77.35
	0.8	88.11	88.37	88.39	87.92	88.05	87.97	86.76
	0.9	94.24	94.39	94.32	94.15	94.55	94.01	93.45
	1	97.70	97.47	97.59	97.50	97.53	97.65	97.38

Figure 2.1: Power Comparison for methods: p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD) and Kernel density procedure (Kernel) and Complete Randomization (CR) in %, simulation based on 10,000 runs and sample size is 100.



From table 2.3, 2.4 and figure 2.1 it is obvious that two sample t-test is less powerful than $lm(z)$ test under these covariate-adaptive designs as well as the complete randomization procedure. Furthermore, compared with complete randomization the covariate-adaptive designs are less powerful when $|\mu_1 - \mu_2|$ is small, but become more powerful when $|\mu_1 - \mu_2|$ is large.

When sample size is small, i.e $N = 32, 64$, in table 2.3 and 2.4, the covariate-adaptive designs have advantage in power under both two sample t-test and covariate test $lm(z)$ which is more obvious in table 2.3 when $N = 32$.

Case 3: Significance of covariates. To investigate type I error of hypothesis testing of the significance of single covariate Z under covariate-adaptive designs and

complete randomization, set the testing hypotheses: $H_0 : \beta_1 = 0$ $v.s.$ $H_a : \beta_1 \neq 0$. Assume the response Y_i is assumed follows linear model:

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + \varepsilon_i$$

where $Z_{i,1} \sim N(0,1)$, $Z_{i,2} \sim exp(1)$ and $\varepsilon_i \sim N(0,1)$. We set $\beta_1 = 0, \beta_2 = 1$, $\mu_1 = \mu_2 = 100$. The simulation is repeated $10,000$ times, biased coin probability $p = 0.8$. The sample size $N = 100, 200, 500$, under significance level $\alpha = 5\%$, testing of the significance of β_1 are performances under tests $lm(z_1), lm(z_1, z_2)$.

Table 2.5: Type I error for $H_0 : \beta_1 = 0$ under methods: p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD) and Kernel density procedure (Kernel) and Complete Randomization (CR) in %, simulation based on 10,000 runs, significance level $\alpha = 5\%$, and sample size $N = 100, 200, 500$.

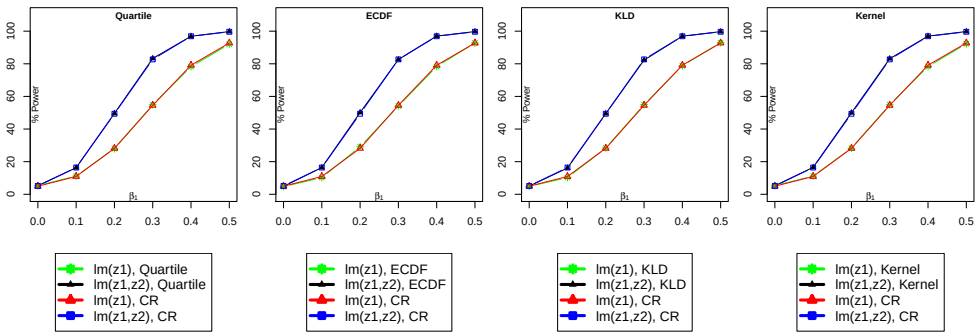
Test	N	P-value	Quartile	ECDF	KLD	Kernel	CR
$lm(z_1)$	100	5.16	5.29	4.72	4.88	5.31	4.95
	200	5.01	4.93	5.04	5.08	4.86	4.87
	500	5.15	5.23	4.76	4.91	5.08	4.77
$lm(z_1, z_2)$	100	5.29	5.13	5.27	5.02	5.22	5.11
	200	4.97	5.09	5.26	4.86	5.17	5.12
	500	5.20	4.82	5.03	4.92	5.09	4.99

Results of simulated type I errors are summarized in table 2.5, from which we can see that both tests are valid in terms of type I error under covariate-adaptive designs and complete randomization procedure.

The comparison of power for different tests is made under restricted random-

izations and complete randomization. Let us set $\beta_1 = 0, 0.1, \cdots, 0.5$, sample size $N = 100$. Simulation results in figure 2.2 shows, the power of testing β_1 is quite similar under all these randomization procedures for the same testing method. $lm(z_1, z_2)$ is more powerful than $lm(z_1)$ under the same randomization procedure.

Figure 2.2: Power Comparison for $H_a : \beta_1 \neq 0$ under methods: Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD) and Kernel density procedure (Kernel) and Complete Randomization (CR) in %, simulation based on 10,000 runs and sample size is 100.



2.5 Appendix: proof of theorems

The following two lemmas are used to the proof of theorem 2.2.1.

Lemma 2.5.1. *Suppose the three conditions in theorem 2.2.1 are satisfied in a covariate-adaptive design, then:*

1. $\sum I_i/N \xrightarrow{P} 1/2$, $\sum(1 - I_i)/N \xrightarrow{P} 1/2$;

2. $\sum I_i X_{i,k}/N \xrightarrow{P} EX_k/2$, $\sum(1 - I_i)X_{i,k}/N \xrightarrow{P} EX_k/2$, $\sum I_i Z_{i,k}/N \xrightarrow{P} EZ_k/2$ *and* $\sum(1 - I_i)Z_{i,k}/N \xrightarrow{P} EZ_k/2$.

Proof. The first part of this lemma says, sample size in both groups will convergence to half of the total patient in probability,

$$\sum I_i/N = \frac{1}{2} + \frac{\sum(2I_i - 1)}{2N} = \frac{1}{2} + o_p(1) \xrightarrow{P} 1/2, \textit{by condition } (1).$$

Same here

$$\sum(1 - I_i)/N = \frac{1}{2} + \frac{(1 - 2I_i)}{2N} = \frac{1}{2} + o_p(1) \xrightarrow{P} 1/2.$$

To prove 2, let us first consider about discrete covariate,

$$\sum I_i X_{i,k}/N = \frac{1}{2N}\sum X_{i,k} + \frac{1}{2N}\sum(2I_i - 1)X_{i,k}.$$

Since for discrete covariate marginal difference $D(k; t_k) = o_p(\sqrt{N})$, and

$$\frac{1}{2N}\sum(2I_i - 1)X_{i,k} = \frac{1}{2N}\sum_{t_k=1}^{s_k} D(k; t_k).$$

Thus $\frac{1}{2N}\sum(2I_i - 1)X_{i,k} = o_p(1)$ by condition (2), and by week law of large number $\frac{1}{2N}\sum X_{i,k} \xrightarrow{P} EX_k/2$, in sum

$$\sum I_i X_{i,k}/N \xrightarrow{P} EX_k/2.$$

When $X_{i,k}$ is continuous covariate, then

$$\sum I_i X_{i,k}/N = \frac{1}{2N}\sum X_{i,k} + \frac{1}{2N}\sum(2I_i - 1)X_{i,k}.$$

Based on condition (3) $D_N^c(X_k) = o_p(N^{1/2})$, it is easy to get $\frac{1}{2N}\sum(2I_i - 1)X_{i,k} =$

$o_p(1)$, and by week law of large number, thus

$$\frac{1}{2N}\sum X_{i,k} \xrightarrow{P} EX_k/2.$$

The proof for Z_i is similar. □

Lemma 2.5.2. *Under the same three conditions in Lemma 2.5.1, $\hat{\alpha}$ is consistent estimator of $\tilde{\alpha}$, which means*

$$\hat{\alpha} \xrightarrow{P} \tilde{\alpha}.$$

Proof. According to the solution of MLE in linear model (1), it is known that

$$\hat{\alpha} = \tilde{\alpha} + (\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\mathbf{Z}}{N}\tilde{\beta} + (\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\tilde{\varepsilon}}{N}.$$

Firstly, we will show $(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\mathbf{Z}}{N}\tilde{\beta} \xrightarrow{P} 0$. By weak law of large number and the independence of each X_i,

$$\frac{1}{N}\mathbf{X}^T\mathbf{X} \xrightarrow{P} diag(\frac{1}{2}, \frac{1}{2}, Var(X_1), ..., Var(X_p)).$$

Since X_i and Z_j are independent, $E(X_iZ_j) = E(X_i)E(Z_j)$ for all i and j; and by

weak law of large number,

$$\frac{1}{N}\mathbf{X}^T\mathbf{Z} = \frac{1}{N}\begin{bmatrix} \sum I_i Z_{i,1} & \dots & \sum I_i Z_{i,q} \\ \sum (1-I_i) Z_{i,1} & \dots & \sum (1-I_i) Z_{i,q} \\ \sum X_{i,1} Z_{i,1} & \dots & \sum X_{i,1} Z_{i,q} \\ \vdots & \ddots & \vdots \\ \sum X_{i,p} Z_{i,1} & \dots & \sum X_{i,p} Z_{i,q} \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \frac{1}{2}EZ_1 & \dots & \frac{1}{2}EZ_q \\ \frac{1}{2}EZ_1 & \dots & \frac{1}{2}EZ_q \\ EX_1EZ_1 & \dots & EX_1EZ_q \\ \vdots & \ddots & \vdots \\ EX_pEZ_1 & \dots & EX_pEZ_q \end{bmatrix}.$$

Since $E(X_k) = E(Z_j) = 0$ for all k and j,

$$(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\mathbf{Z}}{N}\tilde{\beta} \xrightarrow{P} 0. \tag{2.18}$$

Secondly, $E(\varepsilon_i) = 0$, I_i and ε_i are independent for any $i = 1, \dots, N$,

$$\frac{1}{N}\mathbf{X}^T\tilde{\varepsilon} = \frac{1}{N}\begin{bmatrix} \sum I_i\varepsilon_i \\ \sum (1-I_i)\varepsilon_i \\ \sum X_{i,1}\varepsilon_i \\ \vdots \\ \sum X_{i,p}\varepsilon_i \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \frac{1}{2}E\varepsilon_i \\ \frac{1}{2}E\varepsilon_i \\ EX_1E\varepsilon_i \\ \vdots \\ EX_p\varepsilon_i \end{bmatrix}.$$

Thus

$$(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\tilde{\varepsilon}}{N} \xrightarrow{P} 0. \tag{2.19}$$

Hence by (2.18) and (2.19), $\hat{\alpha} \xrightarrow{P} \tilde{\alpha}$. □

Next we will use the above two lemmas to prove Theorem 2.2.1.

Proof of Theorem 2.2.1

Proof. The hypothesis testing of $\mu_1 - \mu_2 = 0$, the test statistic is

$$T = \frac{L\hat{\alpha}}{(\hat{\sigma}^2 L(\mathbf{X}^T\mathbf{X})^{-1}L^T)^{1/2}},$$

where $L = c(1, -1, 0, \ldots, 0)$ and $\hat{\sigma}^2 = (\tilde{Y} - \mathbf{X}\hat{\alpha})^T(\tilde{Y} - \mathbf{X}\hat{\alpha})/(N - p - 2)$, here p is the total number of independent variables besides μ.

First we check the numerator of the test statistic:

$$L\hat{\alpha} = \hat{\mu}_1 - \hat{\mu}_2 = \mu_1 - \mu_2 + L(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}(\frac{\mathbf{X}^T\mathbf{Z}}{N})\tilde{\beta} + L(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\tilde{\varepsilon}}{N}.$$

As show in lemma 2.5.2,

$$\frac{1}{N}\mathbf{X}^T\mathbf{X} \xrightarrow{P} M, \quad where\ M = diag(\frac{1}{2}, \frac{1}{2}, Var(X_1), ..., Var(X_p)).$$

Then the test statistic can be rewritten as:

$$L\hat{\alpha} = \mu_1 - \mu_2 + LM^{-1}(\frac{\mathbf{X}^T\mathbf{Z}}{N}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{N}) + L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})(\frac{\mathbf{X}^T\mathbf{Z}}{N}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{N}).$$

Define $A = LM^{-1}(\frac{\mathbf{X}^T\mathbf{Z}}{N}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{N})$ and $B = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})(\frac{\mathbf{X}^T\mathbf{Z}}{N}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{N})$.
Firstly, it is not hard to find

$$A = \frac{2}{N}(\sum_j \sum_i (2I_i - 1)\beta_j Z_{i,j} + \sum_i (2I_i - 1)\varepsilon_i).$$

Since all margins with respect to each covariate Z_j, the difference between two groups are bounded by condition (2) and (3), that is, $\sum_i (2I_i - 1)\beta_j Z_{i,j} = o_p(N^{1/2})$. For model with finite number of covariates, it follows that

$$\frac{2}{N}(\sum_j \sum_i (2I_i - 1)\beta_j Z_{i,j}) = o_P(N^{-1/2}).$$

Define $\tilde{I} = \{I_i, i = 1, ..., N\}$, and $\tilde{\varepsilon}$ is independent of $\tilde{I}$ given $\mathbf{Z}$, thus $E(2I_i - 1)\varepsilon_i = 0$ and

$$\frac{2}{N} E(\sum_i (2I_i - 1)\varepsilon_i | \mathbf{Z}) = 0,$$

and

$$Var(\frac{2}{N}\sum_i (2I_i - 1)\varepsilon_i | \mathbf{Z}) = \frac{4\sigma_\varepsilon^2}{N}.$$

By the central limit theorem, given $(\tilde{I}, \mathbf{Z})$,

$$\frac{2}{\sqrt{N}}(\sum_i (2I_i - 1)\varepsilon_i | \mathbf{Z}) \xrightarrow{P} N(0, 4\sigma_\varepsilon^2).$$

And, by condition (2) and (3),

$$\frac{2}{\sqrt{N}} \sum_j \sum_i (2I_i - 1)\beta_j Z_{i,j} \xrightarrow{P} 0.$$

By Slutsky theorem,

$$\sqrt{N}A \xrightarrow{D} N(0, 4\sigma_\varepsilon^2). \tag{2.20}$$

Next, we will show $\sqrt{N}B \xrightarrow{P} 0$.

$$\sqrt{N}B = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})(\frac{\mathbf{X}^T\mathbf{Z}}{\sqrt{N}}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{\sqrt{N}}).$$

From Lemma 2.5.2, we have

$$(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1} \xrightarrow{P} 0. \tag{2.21}$$

So to proof $\sqrt{N}B \xrightarrow{P} 0$, it suffices to show

$$\frac{\mathbf{X}^T\mathbf{Z}}{\sqrt{N}}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{\sqrt{N}} = O_p(1).$$

Notice that

$$\frac{\mathbf{X}^T\mathbf{Z}}{\sqrt{N}}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{\sqrt{N}} = \frac{1}{\sqrt{N}}\begin{bmatrix} \sum_j \sum_i I_i Z_{i,j}\beta_i + \sum_i I_i\varepsilon_i \\ \sum_j \sum_i (1-I_i) Z_{i,j}\beta_i + \sum_i (1-I_i)\varepsilon_i \\ \sum_j \sum_i X_{i,1} Z_{i,j}\beta_j + \sum_i X_{i,1}\varepsilon_i \\ \vdots \\ \sum_j \sum_i X_{i,p} Z_{i,j}\beta_j + \sum_i X_{i,p}\varepsilon_i \end{bmatrix}.$$

Also,

$$\begin{aligned} \tfrac{1}{\sqrt{N}}(\textstyle\sum_j \sum_i I_i Z_{i,j}\beta_j + \sum_i I_i\varepsilon_i) &= \tfrac{1}{2}(\tfrac{1}{\sqrt{N}} \textstyle\sum_j \sum_i Z_{i,j}\beta_j + \tfrac{1}{\sqrt{N}} \sum_i \varepsilon_i \\ &\quad + \tfrac{1}{\sqrt{N}} \textstyle\sum_j \sum_i (2I_i - 1) Z_{i,j}\beta_j + \tfrac{1}{\sqrt{N}} \sum_i (2I_i - 1)\varepsilon_i). \end{aligned}$$

By central limit theorem, and the finiteness of covariate number Z_j,

$$\frac{1}{\sqrt{N}} \sum_j \sum_i Z_{i,j}\beta_j + \frac{1}{\sqrt{N}} \sum_i \varepsilon_i = O_p(1). \tag{2.22}$$

Furthermore,

$$\frac{1}{\sqrt{N}}\sum_j\sum_i(2I_i-1)Z_{i,j}\beta_j+\frac{1}{\sqrt{N}}\sum_i(2I_i-1)\varepsilon_i=\frac{\sqrt{N}}{2}A.$$

Since $\frac{\sqrt{N}}{2}A$ converges to a normal distribution,

$$\frac{1}{\sqrt{N}}\sum_j\sum_i(2I_i-1)Z_{i,j}\beta_j+\frac{1}{\sqrt{N}}\sum_i(2I_i-1)\varepsilon_i=O_p(1). \tag{2.23}$$

Hence, based on (2.22) and (2.23)

$$\frac{1}{\sqrt{N}}(\sum_j\sum_i I_iZ_{i,j}\beta_j+\sum_i I_i\varepsilon_i)=O_p(1). \tag{2.24}$$

Also, by central limit theorem, for any $k, k=1,2,...,p$

$$\frac{1}{\sqrt{N}}(\sum_j\sum_i X_{i,k}Z_{i,j}\beta_j+\sum_i X_{i,k}\varepsilon_i)=O_p(1). \tag{2.25}$$

By (2.21), (2.24) and (2.25), we get

$$\sqrt{N}B\xrightarrow{P}0. \tag{2.26}$$

Based on (2.20), (2.26) and Slutsky theorem.

$$\sqrt{N}[\hat{\mu}_1-\hat{\mu}_2-(\mu_1-\mu_2)]\xrightarrow{D}N(0,4\sigma_\varepsilon^2).$$

Up to now, we get the distribution of the numerator part of the test statistic, it is asymptotic normal with mean zero and variance $4\sigma_\varepsilon^2$. To proof the test is conservative, next we will show, the denominator is larger than $4\sigma_\varepsilon^2$. For denominator

part,

$$\hat{\sigma}^2 L(\mathbf{X}^T\mathbf{X})^{-1}L^T.$$

Notice,

$$L(\mathbf{X}^T\mathbf{X})^{-1}L^T = \frac{1}{N}L(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}L^T = \frac{4}{N} + o_p(\frac{1}{N}).$$

Moreover, since $\hat{\alpha}$ is consistent estimator of $\tilde{\alpha}$, we can get the asymptotic variance:

$$\begin{aligned}
\hat{\sigma}^2 &= \tfrac{1}{N-p-2}(\tilde{Y} - \mathbf{X}\hat{\alpha})^T(\tilde{Y} - \mathbf{X}\hat{\alpha}) \\
&= \tfrac{1}{N-p-2}[(\tilde{Y} - \mathbf{X}\tilde{\alpha})^T(\tilde{Y} - \mathbf{X}\tilde{\alpha}) + (\hat{\alpha} - \tilde{\alpha})^T\mathbf{X}^T\mathbf{X}(\hat{\alpha} - \tilde{\alpha}) - 2(\hat{\alpha} - \tilde{\alpha})\mathbf{X}^T(\tilde{Y} - \mathbf{X}\tilde{\alpha})] \\
&= \tfrac{1}{N-p-2}(\mathbf{Z}\tilde{\beta} + \tilde{\varepsilon})^T(\mathbf{Z}\tilde{\beta} + \tilde{\varepsilon}) + o_p(1) \\
&\xrightarrow{P} \sigma_\varepsilon^2 + \textstyle\sum_{j=1}^q \beta_j^2 Var(Z_j).
\end{aligned}$$

Thus, it is easy to get

$$\hat{\sigma}^2 L(\mathbf{X}^T\mathbf{X})^{-1}L^T = \frac{4}{N}(\sigma_\varepsilon^2 + \sum_{j=1}^{q}\beta_j^2 Var(Z_j)) + o_p(\frac{1}{N}).$$

Under $H_0 : \mu_1 - \mu_2 = 0$,

$$T \xrightarrow{D} N(0, \tau^2), \tau^2 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sum_{j=1}^q \beta_j^2 Var(Z_j)} = \frac{\sigma_\varepsilon^2}{\sigma_z^2}$$

for any $\beta_j \neq 0, \sigma_\varepsilon^2 < \sigma_z^2$. When $N \to \infty$

$$Pr(|T| > Z_{(1-\alpha/2)}) \to 2\Phi(-\frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\varepsilon}) < \alpha.$$

Under $H_a : \mu_1 - \mu_2 \neq 0$, $\mu_1 - \mu_2 = \delta/\sqrt{N}$ for fixed $\delta \neq 0$,

$$T \xrightarrow{D} N(\Delta, \tau^2),\ where\ \Delta = \frac{\delta}{2\sigma_z}.$$

Finish the proof of Theorem 2.2.1 □

Proof of theorem 2.2.2

Proof. For the general case of hypothesis testing $H_0 : C\tilde{\alpha} = \xi_0$, the test statistic:

$$T^* = \frac{m^{-1}\sqrt{N}(C\hat{\alpha} - \xi_0)^T[C(\mathbf{X}^T\mathbf{X}/N)^{-1}C^T]^{-1}\sqrt{N}(C\hat{\alpha} - \xi_0)}{\hat{\sigma}^2}.$$

Notice, under $H_0 : C\tilde{\alpha} = \xi_0$,

$$\sqrt{N}(C\hat{\alpha} - \xi_0) = \sqrt{N}C(\hat{\alpha} - \tilde{\alpha}).$$

Notice (2.21) and (2.25), we have

$$\begin{aligned}
\sqrt{N}C(\hat{\alpha} - \tilde{\alpha}) &= \sqrt{N}C(\tfrac{\mathbf{X}^T\mathbf{X}}{N})^{-1}(\tfrac{\mathbf{X}^T\mathbf{Z}\tilde{\beta}}{N} + \tfrac{\mathbf{X}^T\tilde{\varepsilon}}{N}) \\
&= \sqrt{N}CM^{-1}(\tfrac{\mathbf{X}^T\mathbf{Z}\tilde{\beta}}{N} + \tfrac{\mathbf{X}^T\tilde{\varepsilon}}{N}) + \sqrt{N}C((\tfrac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})(\tfrac{\mathbf{X}^T\mathbf{Z}\tilde{\beta}}{N} + \tfrac{\mathbf{X}^T\tilde{\varepsilon}}{N}) \\
&= \sqrt{N}CM^{-1}(\tfrac{\mathbf{X}^T\mathbf{Z}\tilde{\beta}}{N} + \tfrac{\mathbf{X}^T\tilde{\varepsilon}}{N}) + o_p(1).
\end{aligned}$$

Let $C = [0_{m\times 2}, \tilde{C}]$, and $M = diag(1/2, 1/2, \tilde{M})$, we can get

$$\sqrt{N}C(\hat{\alpha} - \tilde{\alpha}) = \tilde{C}\tilde{M}^{-1}\sqrt{N}\begin{bmatrix} \sum_j \sum_i X_{i,1}Z_{i,j}\beta_j + \sum_i X_{i,1}\varepsilon_i \\ \vdots \\ \sum_j \sum_i X_{i,p}Z_{i,j}\beta_j + \sum_i X_{i,p}\varepsilon_i \end{bmatrix} + o_p(1).$$

Then by central limit theorem, and the fact that $CM^{-1}C^T = \tilde{C}\tilde{M}^{-1}\tilde{C}^T$,

$$\sqrt{N}C(\hat{\alpha} - \tilde{\alpha}) \xrightarrow{D} N(0, \sigma_z^2 CM^{-1}C^T).$$

Under $H_0 : C\tilde{\alpha} = \xi_0$,

$$T^* \xrightarrow{D} \chi^2_{(m)}/m.$$

Under $H_a : C\tilde{\alpha} = \xi_1$, and $(\xi_1 - \xi_0) = \eta/\sqrt{N}$, for fixed $\eta \neq 0$,

$$T^* \xrightarrow{D} \chi^2_{(m)}(\lambda)/m, \lambda = \eta^T[CM^{-1}C^T]^{-1}\eta\sigma_z^2.$$

End of proof theorem 2.2.2.

□

Following three Lemmas are used to prove Theorem 2.3.1.

Lemma 2.5.3. $E(\overline{Z}_{n_1}) = 0,\ E(\overline{Z}_{n_2}) = 0.$

Proof. According to the definition, $2I_i - 1$ is the allocation indicator only takes value 1 or -1 with symmetric probability function for two groups, thus $E((2I_i - 1)Z_i) = 0$.

Moreover, $E(Z_i) = 0$, thus

$$\begin{aligned} E(\sum_{i=1}^{n} I_i Z_i) &= 1/2E(\sum Z_i) + 1/2E(\sum(2I_i - 1)Z_i) \\ &= 1/2 \sum EZ_i + 1/2 \sum E((2I_i - 1)Z_i) \\ &= 0. \end{aligned}$$

So it is easy to get $E(\overline{Z}_{n_1}) = 0$, also $E(\overline{Z}_{n_2}) = 0$. □

Lemma 2.5.4. *Define $Z'_n = \frac{n}{2}(\overline{Z}_{n_1} - \overline{Z}_{n_2})$, then a restricted randomization to minimize Z'^2_n is equivalent to the randomization procedure to minimize the mean difference among two groups.*

Proof. For the procedure of minimizing the mean difference among two groups, the imbalance measure is

$$Imb = (\overline{Z}_{n_1} - \overline{Z}_{n_2})^2$$

it is easy to verify,

$$Z'^2_n = Imb * n^2/4.$$

Thus minimizing the mean difference among two groups is equivalent to minimize Z'^2_n. So the randomization procedure using Z'^2_n as imbalance measure is equivalent to the restricted randomization to minimize the mean difference among two groups. □

Lemma 2.5.5. *Define $\Lambda_n = \sum_{i=1}^{n_1} Z_{i1} - \sum_{i=1}^{n_2} Z_{i2}$, and $Z'_n = \frac{n}{2}(\overline{Z}_{n_1} - \overline{Z}_{n_2})$ (same as in Lemma 2.5.4). In addition, we have $\Lambda^{(1)}_{n+1}, Z^{(1)'}_{n+1}$ if $n + 1^{th}$ patient is assigned to group 1 otherwise we have $\Lambda^{(2)}_{n+1}, Z^{(2)'}_{n+2}$ if assigned to group 2. Then,*

(i) $P(\{|\Lambda_n - Z'_n| > \varepsilon \ \ i.o.\}) = 0$, where i.o. means infinitely often;

(ii) $P(|\Lambda_{n+1}^{(1)}| \leq |\Lambda_{n+1}^{(2)}| \& |Z_{n+1}^{(1)'}| > |Z_{n+1}^{(2)'}|,\ i.o.) = 0$*;*
(iii) $P(|\Lambda_{n+1}^{(1)}| > |\Lambda_{n+1}^{(2)}| \& |Z_{n+1}^{(1)'}| \leq |Z_{n+1}^{(2)'}|,\ i.o.) = 0$*;*

Proof. Borel-cantelli Lemma is used to prove these three conclusions. Borel-cantelli Lemma states, for a sequence of random variables $\{X_n\}$, suppose, for any $\varepsilon > 0$, that $A_n(\varepsilon)$ is the event

$$A_n(\varepsilon) \doteq \{w : |X_n(w) - X(w)| > \varepsilon\}$$

if the sum of probabilities of events

$$\sum_{n=1}^{\infty} Pr(A_n(\varepsilon)) = \sum_{n=1}^{\infty} P[|X_n(w) - X(w)| > \varepsilon] < \infty$$

then

$$X_n \xrightarrow{a.s.} X.$$

Indeed, convergences almost surely $\Lambda_n - Z_n' \xrightarrow{a.s.} 0$ implies that for all $\varepsilon > 0$,

$$P(\{|\Lambda_n - Z_n'| > \varepsilon \ \ i.o.\}) = 0.$$

To prove (i) is equivalent to show for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P[|\Lambda_n - Z_n'| > \varepsilon] < \infty. \tag{2.27}$$

Firstly, let us calculate the value of $\Lambda_n - Z'_n$,

$$\begin{aligned} \Lambda_n - Z'_n &= \textstyle\sum_{i=1}^{n_1} Z_{i1} - \sum_{i=1}^{n_2} Z_{i2} - \frac{n}{2}(\overline{z}_{n_1} - \overline{z}_{n_2}) \\ &= (n_1\overline{z}_{n_1} - n_2\overline{z}_{n_2}) - \tfrac{n}{2}(\overline{z}_{n_1} - \overline{z}_{n_2}) \\ &= \tfrac{n_1-n_2}{2} * \overline{Z}_{n_1} + \tfrac{n_1-n_2}{2} * \overline{Z}_{n_2}. \end{aligned} \tag{2.28}$$

We will show that, for any $\varepsilon > 0$,

$$P(|\Lambda_n - Z'_n| > \varepsilon) \leq P(|\frac{\sum_n Z_i}{n}| > \frac{\varepsilon}{2d}) + P(|\frac{d(\overline{Z}_{n_1} - \overline{Z}_{n_2})}{n}| > \frac{\varepsilon}{d}).$$

According to (2.28),

$$P(|\Lambda_n - Z'_n| > \varepsilon) = P(|\frac{n_1 - n_2}{2} * (\overline{Z}_{n_1} + \overline{Z}_{n_2})| > \varepsilon).$$

Since according to the assumption of the randomization procedure that the patient number difference between two groups must be less or equal than d, which means $|n_1 - n_2| \leq d$, thus

$$P(|\Lambda_n - Z'_n| > \varepsilon) \leq P(|\frac{d}{2} * (\overline{Z}_{n_1} + \overline{Z}_{n_2})| > \varepsilon).$$

Also,

$$P(|\Lambda_n - Z'_n| > \varepsilon) \leq P(|\frac{n_1\overline{Z}_{n_1} + n_2\overline{Z}_{n_2}}{n} + \frac{n_2 - n_1}{2n}(\overline{Z}_{n_1} - \overline{Z}_{n_2})| > \frac{\varepsilon}{d}).$$

Moreover,

$$P(|\Lambda_n - Z_n'| > \varepsilon) \leq P(|\frac{(\sum_{n_1} Z_{i1} + \sum_{n_2} Z_{i2})}{n}| + |\frac{n_2 - n_1}{2n}(\overline{Z}_{n_1} - \overline{Z}_{n_2})| > \frac{\varepsilon}{d}).$$

Also, not hard to find

$$P(|\Lambda_n - Z_n'| > \varepsilon) \leq P(|\frac{(\sum_{n_1} Z_{i1} + \sum_{n_2} Z_{i2})}{n}| + |\frac{d}{2n}(\overline{Z}_{n_1} - \overline{Z}_{n_2})| > \frac{\varepsilon}{d}).$$

Moreover, obvioulsy,

$$P(|\Lambda_n - Z_n'| > \varepsilon) \leq P(|\frac{\sum_n Z_i}{n}| > \frac{\varepsilon}{2d}) + P(|\frac{d(\overline{Z}_{n_1} - \overline{Z}_{n_2})}{n}| > \frac{\varepsilon}{d}).$$

To prove (2.27), we only need to show that

$$\sum_{n=1}^{\infty} P(|\frac{\sum_n Z_i}{n}| > \frac{\varepsilon}{2d}) < \infty$$

and

$$\sum_{n=1}^{\infty} P(|\frac{d(\overline{Z}_{n_1} - \overline{Z}_{n_2})}{n}| > \frac{\varepsilon}{d}) < \infty.$$

Firstly, since $Z_1, Z_2, \cdots, Z_n$ are $i.i.d$ with $E(Z_i) = 0$, according to the proof of strong law of large numbers in Etemadi (1981),

$$\sum_{n=1}^{\infty} P(|\frac{\sum_n Z_i}{n}| > \frac{\varepsilon}{2d}) < \infty.$$

Secondly, since $E(\overline{Z}_{n_1} - \overline{Z}_{n_2}) = 0$ and by Chebyshev $P(|\frac{d(\overline{Z}_{n_1} - \overline{Z}_{n_2})}{n}| > \frac{\varepsilon}{d}) \leq \frac{d^4}{n^2\varepsilon^2} *$

$Var(\overline{Z}_{n_1} - \overline{Z}_{n_2})$, moreover, $\frac{1}{n^2}$ is summable, thus

$$\sum_{n=1}^{\infty} P(|\frac{d(\overline{Z}_{n_1} - \overline{Z}_{n_2})}{n}| > \frac{\varepsilon}{d}) < \infty.$$

In sum,

$$\sum_{n=1}^{\infty} P(|\Lambda_n - Z'_n| > \varepsilon) \leq \sum_{n=1}^{\infty} P(|\frac{\sum_n Z_i}{n}| > \frac{\varepsilon}{2d}) + \sum_{n=1}^{\infty} P(|\frac{d(\overline{Z}_{n_1} - \overline{Z}_{n_2})}{n}| > \frac{\varepsilon}{d}) < \infty.$$

By the Borel-cantelli lemma, for any $\varepsilon > 0$,

$$P(\{|\Lambda_n - Z'_n| > \varepsilon \ \ i.o.\}) = 0.$$

Here, we finish the proof of (i). Next we are going to prove (ii) and (iii).

To prove (ii) and (iii), it is enough to prove

$$(|\Lambda^{(1)}_{n+1}| - |\Lambda^{(2)}_{n+1}|) - (|Z^{(1)'}_{n+1}| - |Z^{(2)'}_{n+1}|) \xrightarrow{a.s.} 0.$$

Define $\Delta = |(|\Lambda^{(1)}_{n+1}| - |\Lambda^{(2)}_{n+1}|) - (|Z^{(1)'}_{n+1}| - |Z^{(2)'}_{n+1}|)|$, we still use Borel-cantelli lemma to do the proof. For any $\varepsilon > 0$,

$$\begin{aligned} P(\Delta > \varepsilon) &= P(|(|\Lambda^{(1)}_{n+1}| - |Z^{(1)'}_{n+1}|) - (|\Lambda^{(2)}_{n+1}| - |Z^{(2)'}_{n+1}|)| > \varepsilon) \\ &\leq P(|(|\Lambda^{(1)}_{n+1}| - |Z^{(1)'}_{n+1}|)| > \tfrac{\varepsilon}{2}) + P(|(|\Lambda^{(2)}_{n+1}| - |Z^{(2)'}_{n+1}|)| > \tfrac{\varepsilon}{2}) \\ &\leq P(|\Lambda^{(1)}_{n+1} - Z^{(1)'}_{n+1}| > \tfrac{\varepsilon}{2}) + P(|\Lambda^{(2)}_{n+1} - Z^{(2)'}_{n+1}| > \tfrac{\varepsilon}{2}). \end{aligned}$$

Since $\Lambda_{n+1}^{(1)} - Z_{n+1}^{(1)'} \xrightarrow{a.s} 0$ by (i), thus

$$\sum_{n=1}^{\infty} P(|\Lambda_{n+1}^{(1)} - Z_{n+1}^{(1)'}| > \frac{\varepsilon}{2}) < \infty.$$

Also $\Lambda_{n+1}^{(2)} - Z_{n+1}^{(2)'} \xrightarrow{a.s} 0$ by (i), thus

$$\sum_{n=1}^{\infty} P(|\Lambda_{n+1}^{(2)} - Z_{n+1}^{(2)'}| > \frac{\varepsilon}{2}) < \infty.$$

In sum,

$$P(|(|\Lambda_{n+1}^{(1)}| - |\Lambda_{n+1}^{(2)}|) - (|Z_{n+1}^{(1)'}| - |Z_{n+1}^{(2)'}|)| > \varepsilon) < \infty.$$

Which means

$$P(|\Lambda_{n+1}^{(1)}| \leq |\Lambda_{n+1}^{(2)}| \& |Z_{n+1}^{(1)'}| > |Z_{n+1}^{(2)'}|, i.o) = 0,$$

as well as

$$P(|\Lambda_{n+1}^{(1)}| > |\Lambda_{n+1}^{(2)}| \& |Z_{n+1}^{(1)'}| \leq |Z_{n+1}^{(2)'}|, i.o) = 0.$$

The end of proof for Lemma 2.5.5. □

The proof of Theorem 2.3.1.

Proof. The proof of Theorem 2.3.1 needs two steps. The first step is to show the procedure applying mean difference based method is equivalent to the procedure using $\Lambda_n^2 = (\sum_{i=1}^{n_1}(z_{i1} - \mu) - \sum_{i=1}^{n_2}(z_{i2} - \mu))^2$ as imbalance measure. The second step is to show that, in the procedure using Λ_n^2 as imbalance measure, Λ_n is positive recurrent Markov chain, which means $\Lambda_n = O_p(1)$.

Firstly, Λ_n is an equivalent imbalance measure for mean difference based method. Lemma 2.5.5 shows, only finite steps, the randomization procedures using Λ_n^2 and $Z_n'^2$ will have different preference for assigning new patients. Define $A_i = \{|\Lambda_i^{(1)}| <$

$|\Lambda_i^{(2)}|\}$ and $B_i = \{|Z_i^{(1)'}| < |Z_i^{(2)'}|\}$. Thus there must exist an finite integer N, for all $i \geq N$,

$$P(A_i \cap B_i^c) = 0 \; and \; P(A_i^c \cap B_i) = 0.$$

which indicates for all $i \geq N$,

$$P(\{A_i \cap B_i\} \cup \{A_i^c \cap B_i^c\}) = 1,$$

alternatively,

$$P(\bigcap_{i \geq N} \{A_i \cap B_i\} \cup \{A_i^c \cap B_i^c\}) = 1.$$

So after this N^{th} patient, these two procedures have the same allocation preference for every step with probability 1. Moreover, Lemma 2.5.4 tells us $Z_n'^2$ is an equivalent imbalance measure to mean difference. Thus Λ_n^2 is also equivalent imbalance measure for mean difference based method.

Secondly, Λ_n is positive recurrent Markov chain. Conditions for drift condition under general space: (1) ϕ-irreducible Markov chain; (2) transit probability $\{P(x, \cdot)\}$ is strongly continuous; (3) exists a compact set K, and a non-negative measurable function g on $\mathcal{X}$ that drift condition holds.

Define $\Gamma_n = c(\sum_{i=1}^{N_1} Z_{i1} - \sum_{i=1}^{N_2} Z_{i2}, N_1 - N_2)$ and $\Lambda_n = \sum_{i=1}^{N_1} Z_{i1} - \sum_{i=1}^{N_2} Z_{i2} = l(\Gamma_n)$, thus Λ_n is a linear transform of Γ_n. Obviously, Γ_n is irreducible Markov chain on space $\mathcal{Z}^2$, and $l(\Gamma_n) = \Lambda_n$ is also irreducible Markov chain on $l(\mathcal{Z}^2)$.

Next, prove $\{\Lambda_n\}$ is a ϕ-irreducible Markov chain. According to the definition of ϕ-irreducible Markov chain: $L(x, A) > 0$ for any $\phi(A) > 0$. Here, $L(x, A) = P(\tau_A < \infty) > \int_A P(x, y)dy > 0$ for any $\phi(A) > 0$. Thus $\{\Lambda_n\}$ is a ϕ-irreducible Markov chain on real line R.

Finally, we need to find a compact set K, and non-negative measurable function

g on $\mathcal{X}$ that drift condition holds. The drift condition here means, for some finite b,

$$\int P(x, dy)g(y) \leq g(x) - 1 + b * I(x, K), \qquad x \in \Omega$$

holds.

Assume $E|Z| < \infty$ and $E(Z^2) < \infty$. We choose $g(z) = z^2$, and $K = [-\frac{1+3EZ^2}{2(2p-1)E|Z|} - E|Z|, \ \frac{1+3EZ^2}{2(2p-1)E|Z|} + E|Z|]$,

$$\Lambda_{n+1} = \Lambda_n + (2I_{n+1} - 1) * Z_{n+1}.$$

Case 1: $|n_1 - n_2| < d$,

$$\begin{aligned} E(g(\Lambda_{n+1})|\mathcal{F}_n) - g(\Lambda_n) &= E(\Lambda_n + (2I_{n+1} - 1) * Z_{n+1})^2 - \Lambda_n^2 \\ &= -2(2p - 1)|\Lambda_n| * E|Z| + E(Z^2). \end{aligned}$$

And if $\Lambda_n \notin K$,

$$E(g(\Lambda_{n+1})|\mathcal{F}_n) - g(\Lambda_n) = -2(2p - 1)|\Lambda_n| * E|Z| + E(Z^2) < -1.$$

Else $\Lambda_n \in K$,

$$E(g(\Lambda_{n+1})|\mathcal{F}_n) = \Lambda_n^2 - 2(2p - 1)|\Lambda_n| * E|Z| + E(Z^2) < \infty.$$

In sum, the drift condition holds with a compact set K and non-negative function g here.

Case 2: $n_1 - n_2 = d > 0$, then

$$\Lambda_{n+2} = \Lambda_n - Z_{n+1} + (2I_{n+1} - 1)Z_{n+2},$$

then

$$\begin{aligned} E(g(\Lambda_{n+2})|\mathcal{F}_n) &= E((\Lambda_n - Z_{n+1} + (2I_{n+2} - 1)Z_{n+2})^2|\Lambda_n) \\ &= E((\Lambda_n - Z_{n+1})^2 + Z_{n+2}^2 + 2(2I_{n+2} - 1)Z_{n+2} * (\Lambda_n - Z_{n+1})|\Lambda_n) \\ &= \Lambda_n^2 + EZ_{n+1}^2 + EZ_{n+2}^2 + 2E((2I_{n+2} - 1)Z_{n+2} * (\Lambda_n - Z_{n+1})|\Lambda_n). \end{aligned}$$

Moreover, we will show

$$E((2I_{n+2} - 1)Z_{n+2} * (\Lambda_n - Z_{n+1})) = -(2p - 1)E|Z_{n+2}| * E|\Lambda_n - Z_{n+1}|.$$

The details of the calculation as follows:

$$\begin{aligned} &E((2I_{n+2} - 1)Z_{n+2} * (\Lambda_n - Z_{n+1})) \\ &= \int\int (\Lambda_n - Z_{n+1})(2I_{n+2} - 1)Z_{n+2} f(z_{n+2}) f(z_{n+1}) dZ_{n+2} dZ_{n+1} \\ &= (\int_{-\infty}^{\Lambda_n} + \int_{\Lambda_n}^{\infty}) \int (\Lambda_n - Z_{n+1})(2I_{n+2} - 1)Z_{n+2} dZ_{n+2} dZ_{n+1} \\ &= -(2p - 1)E|Z_{n+2}|(\int_{-\infty}^{\Lambda_n} f(z_{n+1})(\Lambda_n - Z_{n+1}) dZ_{n+1} - \int_{\Lambda_n}^{\infty} f(z_{n+1})(\Lambda_n - Z_{n+1}) dZ_{n+1}) \\ &= -(2p - 1)E|Z_{n+2}| * E|\Lambda_n - Z_{n+1}|. \end{aligned}$$

Thus, in sum

$$\begin{aligned} E(g(\Lambda_{n+2})|\mathcal{F}_n) &= \Lambda_n^2 + 2EZ^2 - 2(2p-1)E|Z_{n+2}| * E|\Lambda_n - Z_{n+1}| \\ &= \Lambda_n^2 + 2EZ^2 - 2(2p-1)E|Z| * E|\Lambda_n - Z| \\ &\leq \Lambda_n^2 + 2EZ^2 - 2(2p-1)E|Z| * (\Lambda_n - E|Z|). \end{aligned}$$

It is easy to verify, when $\Lambda_n \notin K$,

$$E(g(\Lambda_{n+2})|\mathcal{F}_n) \leq g(\Lambda_n) - 1.$$

Else when $\Lambda_n \in K$,

$$E(g(\Lambda_{n+2})|\mathcal{F}_n) \leq \Lambda_n^2 + 2EZ^2 - 2(2p-1)E|Z| * (\Lambda_n - E|Z|) < \infty.$$

We showed drift condition also applies for case 2, and it is the same for the case that $n_1 - n_2 = -d < 0$. In sum, $\{\Lambda_n\}$ is a positive recurrent Markov chain, and it is bounded in probability. Thus Z'_n is also bounded in probability, and $|z_{n_1} - z_{n_2}| = o_p(\frac{1}{\sqrt{n}})$ must hold. □

Chapter 3

Statistical inference for adaptive designs balance correlated continuous covariates

In chapter 2, the statistical inference of adaptive randomized clinical trials based on independent covariates has been studied. The theoretical properties are derived on the assumption that all the covariates in randomization are independent with each other. Based on the independence and linear relationship we have the results that the hypothesis testing to compare treatment effects between two groups is conservative and the testing about a linear combination of covariates remain valid. However, the assumption that all covariates are independent is usually not satisfied in practice. This chapter addresses the problem that what is the theoretical results for hypothesis testing for linear models with correlated covariates. The consistency of the estimators of treatment effects and covariate effects will be studied. The framework to study statistical inference for linear models with correlated covariates is described in the section 3.1. The theoretical properties about hypotheses testing

are given and discussed in section 3.2. Simulations and conclusions are given at last two sections respectively.

3.1 Framework

Here we give the general framework to study statistical inference of linear models with correlated covariates for covariate-adaptive randomized clinical trails. In the general case, the covariates can be correlated by incorporating a covariance matrix into the underlying model. Similar to the independence case in chapter 2, the underlying model and working model are given. The working model only contains partial covariate information used in randomization. The main difference for the dependence cases from the independence case is that the covariance matrix can be any semi-definite matrix instead of diagonal matrix.

There is a covariate-adaptive randomized clinical trial, with two treatments: 1 and 2. The discrete and continuous covariates of patients are balanced by applying one covariate-adaptive design. Let μ_j, $j = 1, 2$ be the expected main effect of treatment j respectively; N be the total number of patients in this trial; I_i, $i = 1, \cdots, N$ be the allocation indicator of patient i, say $I_i = 1$ if patient i is assigned to treatment 1, $I_i = 0$ otherwise. Assume the relationship between the response Y and covariates, treatment follows the linear model below:

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + \alpha_1 X_{i,1} + \cdots + \alpha_p X_{i,p} + \beta_1 Z_{i,1} + \cdots + \beta_q Z_{i,q} + \varepsilon_i \quad (3.1)$$

where

1. $X_{i,k}$, $k = 1, \cdots, p$ is discrete or continuous covariate identically independent distributed with as X_k, with $E(X_k) = 0$, which is used in both covariate

randomization procedure and final statistical inference.

2. $Z_{i,j}$, $j = 1, \cdots, q$ is discrete or continuous covariate identically independent distributed with as Z_j, with $E(Z_j) = 0$, which is only used in covariate randomization procedure.

3. ε_i, $i = 1, \cdots, N$ is independent and identically distributed random error with $E(\varepsilon_i) = \sigma_\varepsilon^2$.

4. $X_{i,k}$, $k = 1, \cdots, p$ and $Z_{i,j}$, $j = 1, \cdots, q$ can be correlated with each other, let the covariance matrix of $(\mathbf{X}, \mathbf{Z})$ be Σ, where

$$\Sigma = \begin{bmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_p) & Cov(X_1, Z_1) & \cdots & Cov(X_1, Z_q) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & \cdots & Cov(X_p, X_p) & Cov(X_p, Z_1) & \cdots & Cov(X_p, Z_q) \\ Cov(Z_1, X_1) & \cdots & Cov(Z_1, X_p) & Cov(Z_1, Z_1) & \cdots & Cov(Z_1, Z_q) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Cov(Z_q, X_1) & \cdots & Cov(Z_q, X_p) & Cov(Z_q, Z_1) & \cdots & Cov(Z_q, Z_q) \end{bmatrix}.$$

5. ε_i is independent with $X_{i,k}$ and $Z_{i,j}$.

Define $\tilde{Y} = (Y_1, Y_2, \cdots, Y_N)^T$, $\tilde{\alpha} = (\mu_1, \mu_2, \alpha_1, \cdots, \alpha_p)^T$, $\tilde{\beta} = (\beta_1, \beta_2, \cdots, \beta_q)^T$ and

$\tilde{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N)^T$,

$$\mathbf{X} = \begin{bmatrix} I_1 & (1-I_1) & X_{1,1} & \cdots & X_{1,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I_N & (1-I_N) & X_{N,1} & \cdots & X_{N,p} \end{bmatrix}, \ \mathbf{Z} = \begin{bmatrix} Z_{1,1} & \cdots & Z_{1,q} \\ \vdots & \ddots & \vdots \\ Z_{N,1} & \cdots & Z_{N,q} \end{bmatrix}$$

$$\Sigma_{x,x} = \begin{bmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_p) \\ \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & \cdots & Cov(X_p, X_p) \end{bmatrix} \ \Sigma_{z,z} = \begin{bmatrix} Cov(Z_1, Z_1) & \cdots & Cov(Z_1, Z_q) \\ \vdots & \ddots & \vdots \\ Cov(Z_q, Z_1) & \cdots & Cov(Z_q, Z_q) \end{bmatrix}$$

$$\Sigma_{x,z} = \Sigma_{z,x}^T = \begin{bmatrix} Cov(X_1, Z_1) & \cdots & Cov(X_1, Z_q) \\ \vdots & \ddots & \vdots \\ Cov(X_p, Z_1) & \cdots & Cov(X_p, Z_q) \end{bmatrix}.$$

The general model used for randomization can be rewritten as,

$$\tilde{Y} = \mathbf{X}\tilde{\alpha} + \mathbf{Z}\tilde{\beta} + \tilde{\varepsilon}.$$

The statistical inference working model would be,

$$\tilde{Y} = \mathbf{X}\tilde{\alpha} + \tilde{\varepsilon}$$

and the expectation of Y_i is,

$$E[Y_i] = \mu_1 I_i + \mu_2(1 - I_i) + \alpha_1 X_{i,1} + \cdots + \alpha_p X_{i,p}. \tag{3.2}$$

REMARK 3.1.1. *Both the underlying model and working model in chapter 3 is similar to those of independence case in chapter 2. The only difference is that the covariance structure of covariates in chapter 3 can be any semi-definite matrix instead of diagonal matrix. In this general case where covariates are not necessarily independent. Besides, both* $X_{i,k}$, $k = 1, \cdots, p$ *and* $Z_{i,j}$, $j = 1, \cdots, q$ *are assumed to be scalars here. For covariates with two categories, a dummy variable is used. In cases with more than two categories, for example when there are three categories, high dimensional vectors (0,0), (0,1), and (1,0) are coded in the model. All the results can be extended to the situation that discrete covariates have multiple categories easily.*

Based on the final statistical inference model (3.2) to compare if there is difference between two treatment effects, do the following hypothesis test:

$$H_o : \mu_1 = \mu_2 \; vs. \; H_a : \mu_1 \neq \mu_2. \tag{3.3}$$

According to the ordinary least square estimation method, the estimate of $\tilde{\alpha}$ is obtained by the following formula,

$$\hat{\alpha} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\tilde{Y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\tilde{\alpha} + \mathbf{Z}\tilde{\beta}).$$

The test statistic for 3.3 is

$$T = \frac{L\hat{\alpha}}{(\hat{\sigma}^2 L(\mathbf{X}^T\mathbf{X})^{-1}L^T)^{1/2}} \tag{3.4}$$

where $L = (1, -1, 0, \cdots, 0)$ and $\hat{\sigma}^2 = (\tilde{Y} - \mathbf{X}\hat{\alpha})^T(\tilde{Y}\hat{\alpha})/(N - p - 2)$. The null hypothesis is rejected when $|T| > Z_{1-\alpha/2}$, and otherwise accepted when $Z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ percentile of the standard normal distribution.

3.2 Theoretical properties

For clinical trials designed to balance both discrete and continuous covariates with covariate-adaptive designs. Suppose discrete and continuous covariates: $X_{i,k}$, $k = 1, \cdots, p$, $Z_{i,j}$, $j = 1, \cdots, q$ are balanced in randomization procedure. Let $D = (k|X_k$ is categorical, $k = 1, ..., p)$, $C = (k|X_k$ is continuous, $k = 1, ..., p)$, $D^* = (j|Z_j$ is categorical, $j = 1, ..., q)$ and $C^* = (j|Z_j$ is continuous, $j = 1, ..., q)$. The marginal imbalance measure for all levels of categorical covariates are considered here. Assume categorical covariate $X_k \in D$ has level s_k and $Z_j \in D$ has level s_j^*. Continuous covariate $X_k \in C$, $Z_k \in C^*$ two group difference is define as the difference of the sum of this covariate in two groups. For i^{th} patient, we use $W_i = (X_{i,1}, X_{i,2}, ..., X_{i,p}, Z_{i,1}, Z_{i,2}, ..., Z_{i,q})$ to represent the covariate profile. If $X_{i,k}$ is categorical and at level t_k $(1 \leq t_k \leq s_k)$ and $Z_{i,j}$ is categorical and at level t_j $(1 \leq t_j \leq s_j^*)$, for convenience, we use $(k; t_k)$ to denote the margin at categorical level $X_k = t_k$, and t_j to denote the margin at categorical level $Z_j = t_j$. The overall and marginal imbalance between two treatments are defined as:

1. D_N be the difference between the number of patients among two groups as total, where $D_N = N_1 - N_2$;

2. $D_N(k; t_k)$, $D_N(j; t_j)$ be the differences between the number of patients in the two treatment groups on the margin $(k; t_k)$ for categorical covariates, where

$D_N(k;t_k) = N_{t_k,1} - N_{t_k,2}$, $D_N(j;t_j) = N_{t_j,1} - N_{t_j,2}$;

3. $D_N^c(X_k)$, $D_N^c(Z_j)$be the difference between continuous covariate k and j among two groups, where $D_N^c(X_k) = \sum_{i_1=1}^{N_1} X_{i_1,k} - \sum_{i_2=1}^{N_2} X_{i_2,k}$, $D_N^c(Z_j) = \sum_{i_1=1}^{N_1} Z_{i_1,j} - \sum_{i_2=1}^{N_2} Z_{i_2,j}$.

REMARK 3.2.1. *Only the hypothesis testing for comparing treatment effects is discussed in chapter 3. Because it can be shown that, the hypothesis testing about treatment effects under complete randomization is still valid, however, the estimators for coefficients of covariate are biased if any covariate omitted from the inference model. The theoretical properties of statistical inference under covariate-adaptive designs satisfy above three conditions are studied in the following theorems. In theorem 3.2.1, the consistency and biases will be given for the estimators of treatment effects and covariate effects. Moreover, theorem 3.2.2 shows the theoretical properties of the hypothesis testing of treatment effects under covariate-adaptive deigns.*

Theorem 3.2.1. *Suppose the following three conditions are satisfied in a covariate-adaptive design:*

(1) the overall imbalance converges to zero in probability by rate $N^{1/2}$, that is $D_N = o_p(N^{1/2})$;

(2) the marginal imbalance for each categorical covariate converges to zero in probability by rate $N^{1/2}$, that is, $D_N(k,t_k) = o_p(N^{1/2})$ and $D_N(j,t_j) = o_p(N^{1/2})$;

(3) each continuous covariate sum in two groups converges to zero in probability by rate $N^{1/2}$, that is, $D_N^c(X_k) = o_p(N^{1/2})$ and $D_N^c(Z_j) = o_p(N^{1/2})$.

(i) The estimator of treatment effect difference is consistent, i.e.,

$$\hat{\mu}_1 - \hat{\mu}_2 \xrightarrow{P} \mu_1 - \mu_2.$$

(ii) $(\hat{\alpha}_1, \cdots, \hat{\alpha}_p)^T$ *is a biased estimator for* $(\alpha_1, \cdots, \alpha_p)^T$*, i.e.,*

$$(\hat{\alpha}_1, \cdots, \hat{\alpha}_p)^T - (\alpha_1, \cdots, \alpha_p)^T \xrightarrow{P} \Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta}.$$

Theorem 3.2.1, gives the theoretical properties of test statistic for testing hypothesis of treatment effects under covariate-adaptive designs. Under the assumption of three mild conditions of covariate-adaptive designs, theorem 3.2.1 shows the difference of estimators of treatment effects is consistent, similar to the independence case in chapter 2. Even if the covariance matrix structure is not fully understood, we can always achieve consistent estimator of treatment effect. On the other hand, it also shows the estimators of covariates coefficients are biased if important covariates are omitted from the inference model. From the results in theorem 3.2.1, we know the covariance matrix between all covariates need to be fully understood to get the exact value of bias. So for some studies where we need to identify the importance of covariates or biomarkers, we need to take care of the bias. The bias of estimators for covariates coefficients exists no matter under covariate-adaptive designs or complete randomization.

Theorem 3.2.2. *Suppose the following three conditions are satisfied in a covariate-adaptive design:*

(1) the overall imbalance converges to zero in probability by rate $N^{1/2}$*, that is* $D_N = o_p(N^{1/2})$;

(2) the marginal imbalance for each categorical covariate converges to zero in probability by rate $N^{1/2}$*, that is,* $D_N(k, t_k) = o_p(N^{1/2})$ *and* $D_N(j, t_j) = o_p(N^{1/2})$*;*

(3) each continuous covariate sum in two groups converges to zero in probability by rate $N^{1/2}$*, that is,* $D_N^c(X_k) = o_p(N^{1/2})$ *and* $D_N^c(Z_j) = o_p(N^{1/2})$.

(i) Then under $H_0 : \mu_1 - \mu_2 = 0$,

$$T \xrightarrow{D} N(0, \tau^2) \text{ , where } \tau^2 = \frac{\sigma_\varepsilon^2}{\sigma_z^2},\ \sigma_z^2 = \sigma_\varepsilon^2 + Var[(\mathbf{Z}^T - \mathbf{X}^T \Sigma_{x,x}^{-1} \Sigma_{x,z})\tilde{\beta}] \tag{3.5}$$

if $Var[(\mathbf{Z}^T - \mathbf{X}^T \Sigma_{x,x}^{-1} \Sigma_{x,z})\tilde{\beta}] = 0$, *then* $\tau^2 = 1$.

(ii) Under $H_a : \mu_1 - \mu_2 \neq 0$, *consider a sequence of local alternative, i.e,* $\mu_1 - \mu_2 = \delta/\sqrt{N}$ *for a fixed* $\delta \neq 0$, *then*

$$T \xrightarrow{D} N(\Delta, \tau^2), \quad where \quad \Delta = \frac{\delta}{2\sigma_z}. \tag{3.6}$$

REMARK 3.2.2. *Under null hypothesis, type I error is conservative when* $\tau^2 < 1$. $\tau^2 = 1$ *only if* $Var[(\mathbf{Z}^T - \mathbf{X}^T \Sigma_{x,x}^{-1} \Sigma_{x,z})\tilde{\beta}] = 0$. *When* $\beta_j = 0$ *for* $j = 1, ..., q$, *all the covariates used in randomization are not related with outcome* Y, *then* $\tau^2 = 1$. *Moreover, if* $(\mathbf{Z}^T - \mathbf{X}^T \Sigma_{x,x}^{-1} \Sigma_{x,z}) = 0$ *which means* $\mathbf{Z}$ *is a linear transform of* $\mathbf{X}$, *then also* $\tau^2 = 1$. *Under these two cases, the hypothesis testing can achieve type I error. Otherwise, the hypothesis testing is conservative.*

REMARK 3.2.3. *Under alternative hypothesis, the power can be obtained from the asymptotic distribution 3.6. Under covariate-adaptive design, power is*

$$P(|T| > Z_{1-\alpha/2}) = \Phi(\frac{\delta}{2\sigma_\varepsilon} - \frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\varepsilon}) + \Phi(-\frac{\delta}{2\sigma_\varepsilon} - \frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\varepsilon}) + o(1).$$

The power of complete randomization would be

$$P(|T| > Z_{1-\alpha/2}) = \Phi(\frac{\delta}{2\sigma_\varepsilon} - Z_{1-\alpha/2}) + \Phi(-\frac{\delta}{2\sigma_\varepsilon} - Z_{1-\alpha/2}) + o(1),$$

because $\sigma_\varepsilon = \sigma_z$ *under complete randomization.*

Some conclusions about the power comparisons of covariate-adaptive designs and complete randomization can be made. First, the asymptotic power under covariate-adaptive design is smaller than complete randomization when δ is small, and usually larger than complete randomization when δ is large. The simulation results also confirm these conclusions.

3.3 Simulation study

3.3.1 Case 1: Testing treatment effects

Continuous covariate randomization methods: p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD) and Kernel density procedure (Kernel) are compared with complete randomization (CR) to compare the type I error of the hypothesis testing for comparing treatment effects. (The details of above continuous covariate-adaptive designs can be found in section 1.2.2)

The response Y_i is assumed to follow the linear model:

$$Y_i = \mu_1 I_i + \mu_2(1 - I_i) + \beta_1 Z_{i,1} + \beta_2 Z_{i,2} + \varepsilon_i \tag{3.7}$$

where $\binom{Z_1}{Z_2} \sim N(\mu, \Sigma)$, $\mu = \binom{0}{0}$, $\Sigma = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$, and $\varepsilon_i \sim N(0,1)$, $\beta_1 = 1,\ \beta_2 = 1$. To investigate the type I error of the hypothesis testing: $H_o :\ \mu_1 = \mu_2$, no treatment effect difference is assumed here, i.e., $\mu_1 = \mu_2$. In the simulation, similar settings of parameters are used here as in chapter 2. Biased coin probability $p = 0.8$ is applied for all the covariate-adaptive designs. The significance level $\alpha = 0.05$ and sample size $N = 100, 200, 500$ is considered.

Table 3.1: Type I error for methods: p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR) in % with $\rho = 0.5$, simulation based on 10,000 runs.

Method	N	t-test	$lm(Z_1)$	$lm(Z_2)$	$lm(Z_1, Z_2)$	B-test
P-value based	100	0.04	1.30	1.31	5.25	4.74
P-value based	200	0.02	1.05	0.99	4.73	4.65
P-value based	500	0.01	0.99	0.95	5.11	5.36
Quartile	100	0.36	2.22	2.25	4.97	5.23
Quartile	200	0.25	2.40	2.19	5.09	4.59
Quartile	500	0.18	2.15	1.90	5.04	4.53
ECDF	100	0.07	1.21	1.37	5.16	5.27
ECDF	200	0.02	1.29	1.30	5.27	5.39
ECDF	500	0.01	1.01	1.03	4.88	5.18
KLD	100	0.06	1.50	1.35	5.26	5.39
KLD	200	0.05	1.16	1.05	5.26	4.52
KLD	500	0.02	1.10	1.01	5.42	4.63
Kernel	100	0.19	2.19	2.12	4.80	5.25
Kernel	200	0.10	1.66	1.80	4.77	4.69
Kernel	500	0.03	1.53	1.30	4.91	5.34
CR	100	5.33	5.17	4.95	4.66	-
CR	200	4.98	5.10	4.75	4.93	-
CR	500	5.12	5.08	5.28	4.89	-

Three types of tests are compared in table 3.1: (1) two sample t-test (t-test);(2) covariate test based on the linear model contains covariate Z $(lm(z), lm(z_1, z_2))$; (3) bootstrap t-test introduced in Shao, Yu, and Zhong (2010), where bootstrap samples $(Y_1^{*b}, Z_1^{*b}), \cdots, (Y_N^{*b}, Z_N^{*b})$, $b = 1, 2, \cdots, B$, are generated independently randomly with replacement from sample $(Y_1, Z_1), \cdots, (Y_N, Z_N)$. The variance of $\bar{Y}_1 - \bar{Y}_2$ is estimated by the bootstrap samples. Based on Table 3.1, several conclusions can

be made: (1) Two sample t-test has the most conservative results with the smallest type I error. (2) Under covariate-adaptive design, $lm(Z_1)$ and $lm(Z_2)$ are also conservative with type I error less than 5%. (3) The linear model with both covariates Z_1 and Z_2, $(lm(Z_1, Z_2))$ is valid for the type I error is close to 5%. These findings are consistent to Theorem 3.2.2 with regards to the hypothesis testing with full model or omitted covariates.

Table 3.2: Type I error for methods: p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR) in % with different values of ρ, simulation based on 10,000 runs, sample size $N = 100$.

Method	ρ	t-test	$lm(Z_1)$	$lm(Z_2)$	$lm(Z_1, Z_2)$	B-test
P-value based	0.8	0.05	2.38	2.49	5.03	4.61
	0.5	0.02	1.34	1.10	4.89	4.84
	0.2	0.08	0.68	0.67	4.75	5.56
	0	0.09	0.64	0.72	5.00	4.92
	-0.2	0.27	0.72	0.65	4.90	5.12
	-0.5	0.63	0.94	1.02	5.00	4.73
	-0.8	2.34	2.49	2.41	4.82	4.87
Quartile	0.8	0.17	3.87	3.71	4.92	5.39
	0.5	0.17	2.27	2.19	4.93	4.60
	0.2	0.63	1.70	1.74	5.23	5.06
	0	0.85	1.50	1.76	4.88	4.80
	-0.2	1.22	1.51	1.52	4.42	4.32
	-0.5	2.17	2.40	2.52	5.14	5.74
	-0.8	3.98	3.86	3.96	4.75	4.56
ECDF	0.8	0.03	2.75	2.80	4.66	4.54
	0.5	0.06	1.28	1.27	5.09	4.59
	0.2	0.03	0.76	0.87	5.05	4.92
	0	0.21	0.87	0.94	5.23	5.41
	-0.2	0.33	0.91	0.88	4.86	5.35
	-0.5	1.07	1.40	1.39	4.91	4.52
	-0.8	2.84	2.82	2.90	5.11	5.61

KLD	0.8	0.01	2.92	2.82	4.91	5.28
	0.5	0.07	1.29	1.18	4.84	5.63
	0.2	0.04	0.77	0.82	0.82	4.54
	0	0.18	0.71	0.67	4.97	5.52
	-0.2	0.35	0.89	0.77	4.89	5.16
	-0.5	0.86	1.20	1.23	4.71	4.90
	-0.8	2.94	3.02	2.92	5.06	4.58
Kernel	0.8	0.08	3.63	3.91	5.26	4.80
	0.5	0.11	1.93	2.19	5.05	5.12
	0.2	0.40	1.32	1.45	4.79	5.09
	0	0.84	1.57	1.58	5.21	5.68
	-0.2	0.90	1.45	1.22	5.11	5.34
	-0.5	2.07	2.25	2.17	5.09	4.73
	-0.8	3.86	3.75	3.90	4.83	4.84
CR	0.8	5.30	5.05	4.96	4.66	-
	0.5	4.89	5.11	4.72	4.64	-
	0.2	4.85	5.04	5.09	5.08	-
	0	4.63	4.91	4.86	5.29	-
	-0.2	4.89	4.61	4.88	4.71	-
	-0.5	5.12	5.21	4.96	5.12	-
	-0.8	5.24	5.35	5.00	4.98	-

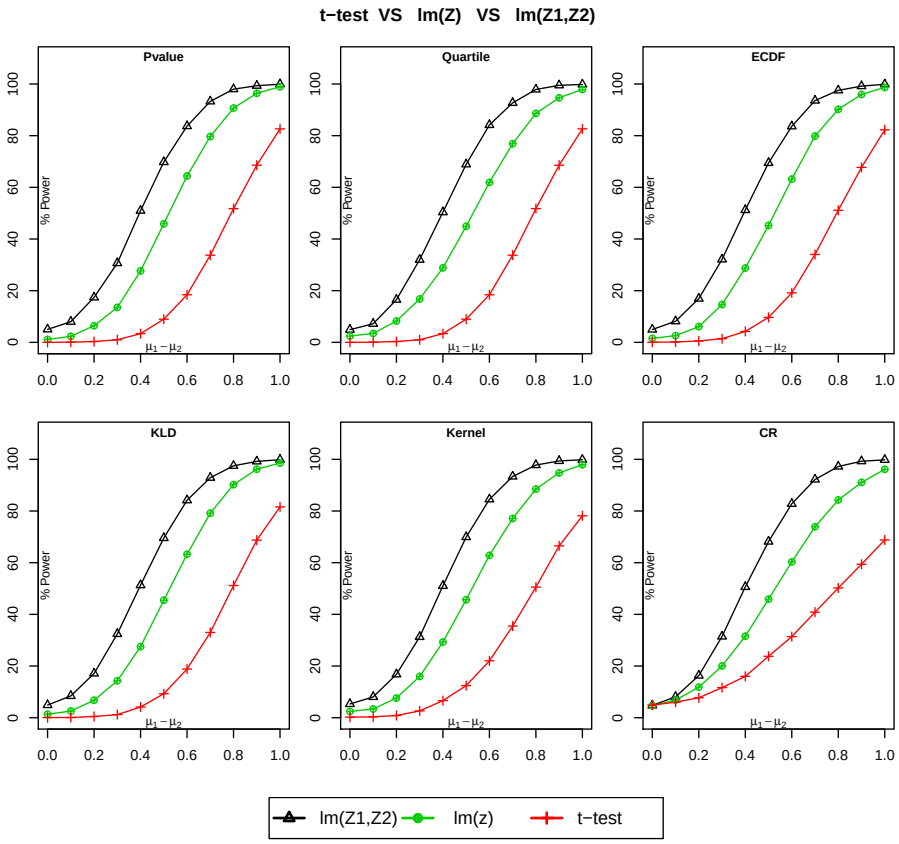
For the case of different correlations between two covariates, i.e., $\rho = (\pm 0.8, \pm 0.5, \pm 0.2, 0)$, the type I error results are similar to Table 3.1 which is also consistent with Theorem 3.2.2.

3.3.2 Case 2: Power comparison

To compare the power of testing treatment effects under continuous covariate-adaptive designs and complete randomization, the same model is used as in case

1. The difference between treatment effects μ_1 and μ_2, i.e., $\mu_1 - \mu_2 \neq 0$. Sample size $N = 100$, $\rho = 0.5$, and the simulation is based on $10,000$ iterations.

Figure 3.1: Power Comparison for methods: p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR) in % with $\rho = 0.5$, simulation based on 10,000 runs and sample size $N = 100$.



Two sample t-test is compared with hypothesis testing including covariates under complete randomization and covariate-adaptive randomizations. Figure 3.1 shows that two sample t-test is less powerful than $lm(z)$ test under covariate-adaptive designs as well as the complete randomization procedure. Furthermore, compared with complete randomization the covariate-adaptive designs are less powerful when

$|\mu_1 - \mu_2|$ is small, but become more powerful when $|\mu_1 - \mu_2|$ increases.

3.4 Conclusion

Chapter 3 discusses the hypothesis testing problem under linear models with correlated covariates. The theoretical properties of statistical inference are derived under the assumption that the covariates are correlated with each other instead of they are independent. Based on the asymptotic distributions of the test statistic, Theorem 3.3.1 shows that hypothesis testing for comparing treatment effects is always conservative. The type I error is always smaller than the given significant level when any important covariates are missed from the inference model. Moreover, two sample t-test is the most conservative among the tests we compared here since it does not use any covariate information. These conclusions are similar with our findings in Chapter 2.

The estimator of covariate coefficient is biased if the covariates are correlated with each other and important covariates are omitted from the analysis model. Thus the hypothesis testing about covariate effect is not valid under the case discussed in Chapter 3. However, the treatment effect estimator is still valid both under the independence case and the dependence case. In realistic, we should be more careful when we consider about the inference of covariate effect because usually we may not be aware of the covariates if they are independent or not.

3.5 Appendix: proof of theorems

Lemma 3.5.1. *Suppose the following three conditions are satisfied in a covariate-adaptive design:*

(1) the overall imbalance converges to zero in probability by rate $N^{1/2}$, that is $D_N = o_p(N^{1/2})$;

(2) the marginal imbalance for each categorical covariate converges to zero in probability by rate $N^{1/2}$, that is, $D_N(k, t_k) = o_p(N^{1/2})$ and $D_N(j, t_j) = o_p(N^{1/2})$;

(3) each continuous covariate sum in two groups converges to zero in probability by rate $N^{1/2}$, that is, $D_N^c(X_k) = o_p(N^{1/2})$ and $D_N^c(Z_j) = o_p(N^{1/2})$.

then

1. $\sum I_i/N \xrightarrow{P} 1/2$,$\sum(1 - I_i)/N \xrightarrow{P} 1/2$;
2. $\sum I_i X_{i,k}/N \xrightarrow{P} EX_k/2$, $\sum(1 - I_i)X_{i,k}/N \xrightarrow{P} EX_k/2$, $\sum I_i Z_{i,k}/N \xrightarrow{P} EZ_k/2$ *and* $\sum(1 - I_i)Z_{i,k}/N \xrightarrow{P} EZ_k/2$.

Proof. The proof is the same as the proof for Lemma 2.5.1. □

Proof of Theorem 3.2.1. According to the solution of OLS in linear model 3.1, it is known that

$$\hat{\alpha} = \tilde{\alpha} + (\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X^T Z}}{N}\tilde{\beta} + (\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\tilde{\varepsilon}}{N}.$$

First, we will show $(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X^T Z}}{N}\tilde{\beta} \xrightarrow{P} 0$.

By weak law of large number,

$$\begin{aligned}
\tfrac{1}{N}\mathbf{X}^T\mathbf{X} &= \tfrac{1}{N}\begin{bmatrix}
\sum I_i & 0 & \sum I_i X_{i,1} & \dots & \sum I_i X_{i,p} \\
0 & (1-\sum I_i) & \sum (1-I_i) X_{i,1} & \dots & \sum (1-I_i) X_{i,p} \\
\sum X_{i,1} I_i & \sum X_{i,1}(1-I_i) & \sum X_{i,1} X_{i,1} & \dots & \sum X_{i,1} X_{i,p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sum X_{i,p} I_i & \sum X_{i,p}(1-I_i) & \sum X_{i,p} X_{i,1} & \dots & \sum X_{i,p} X_{i,p}
\end{bmatrix} \\
&\xrightarrow{P} \; diag(\tfrac{1}{2}, \tfrac{1}{2}, \Sigma_{x,x})
\end{aligned}$$

and by weak law of large number,

$$\frac{1}{N}\mathbf{X^T Z} = \frac{1}{N}\begin{bmatrix}
\sum I_i Z_{i,1} & \dots & \sum I_i Z_{i,q} \\
\sum (1-I_i) Z_{i,1} & \dots & \sum (1-I_i) Z_{i,q} \\
\sum X_{i,1} Z_{i,1} & \dots & \sum X_{i,1} Z_{i,q} \\
\vdots & \ddots & \vdots \\
\sum X_{i,p} Z_{i,1} & \dots & \sum X_{i,p} Z_{i,q}
\end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0 \\ 0 \\ \Sigma_{x,z} \end{bmatrix}$$

for $E(X_k) = E(Z_j) = 0$ for all k and j, and by the weak law of large numbers and dependence of covariates X and Z.

Secondly, $E(\varepsilon_i) = 0$, I_i and ε_i are independent for any $i = 1, \ldots, N$,

$$\frac{1}{N}\mathbf{X}^T\tilde{\varepsilon} = \frac{1}{N}\begin{bmatrix} \sum I_i\varepsilon_i \\ \sum(1-I_i)\varepsilon_i \\ \sum X_{i,1}\varepsilon_i \\ \vdots \\ \sum X_{i,p}\varepsilon_i \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \frac{1}{2}E\varepsilon_i \\ \frac{1}{2}E\varepsilon_i \\ EX_1\varepsilon_i \\ \vdots \\ EX_p\varepsilon_i \end{bmatrix}.$$

Thus

$$(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X^T Z}}{N}\tilde{\beta} \xrightarrow{P} \begin{bmatrix} 0 \\ 0 \\ \Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta} \end{bmatrix}$$

and

$$(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\tilde{\varepsilon}}{N} \xrightarrow{P} 0.$$

Hence,

$$(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X^T Z}}{N}\tilde{\beta} + (\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\tilde{\varepsilon}}{N} \xrightarrow{P} \begin{bmatrix} 0 \\ 0 \\ \Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta} \end{bmatrix}$$

so,

$$\hat{\alpha} - \tilde{\alpha} \xrightarrow{P} \begin{bmatrix} 0 \\ 0 \\ \Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta} \end{bmatrix}.$$

Thus,

$$\hat{\mu}_1 - \hat{\mu}_2 \xrightarrow{P} \mu_1 - \mu_2$$

moreover,

$$(\hat{\alpha}_1, \ldots, \hat{\alpha}_p)^T - (\alpha_1, \ldots, \alpha_p)^T \xrightarrow{P} \Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta}.$$

□

Proof of Theorem 3.2.2. For hypothesis testing of $\mu_1 - \mu_2 = 0$, the test statistic is

$$T = \frac{L\hat{\alpha}}{(\hat{\sigma}^2 L(\mathbf{X}^T\mathbf{X})^{-1}L^T)^{1/2}}$$

where $L = (1, -1, 0, \ldots, 0)$ and $\hat{\sigma}^2 = (\tilde{Y} - \mathbf{X}\hat{\alpha})^T(\tilde{Y} - \mathbf{X}\hat{\alpha})/(N - p - 2)$, here p is the total number of independent variables besides μ.

First have a look at the numerator of the test statistic:

$$L\hat{\alpha} = \hat{\mu}_1 - \hat{\mu}_2 = \mu_1 - \mu_2 + L(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}(\frac{\mathbf{X}^\mathbf{T}\mathbf{Z}}{N})\tilde{\beta} + L(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}\frac{\mathbf{X}^T\tilde{\varepsilon}}{N}.$$

As show in lemma 2.5.2,

$$\frac{1}{N}\mathbf{X}^T\mathbf{X} \xrightarrow{P} diag(\frac{1}{2}, \frac{1}{2}, \Sigma_{x,x}) \doteq M.$$

Then the test statistic can be rewritten as:

$$L\hat{\alpha} = \mu_1 - \mu_2 + LM^{-1}(\frac{\mathbf{X^T Z}}{N}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{N}) + L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})(\frac{\mathbf{X^T Z}}{N}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{N}).$$

Define

$$A = LM^{-1}[\frac{\mathbf{X^T Z}}{N}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{N}]$$

and

$$B = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})(\frac{\mathbf{X^T Z}}{N}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{N}).$$

Firstly, after some calculation

$$A = \frac{2}{N}(\sum_j \sum_i (2I_i - 1)\beta_j Z_{i,j} + \sum_i (2I_i - 1)\varepsilon_i).$$

Since all margins with respect to each covariate Z_j, the difference between two groups are bounded by condition (2) and (3), it is, $\sum_i (2I_i - 1)\beta_j Z_{i,j} = o_P(N^{1/2})$. For model with limited covariates, it follows that

$$\frac{2}{N}(\sum_j \sum_i (2I_i - 1)\beta_j Z_{i,j}) = o_P(N^{-1/2}).$$

Define $\mathbf{Z} = \{Z_{i,j}, i = 1, 2, ..., N, j = 1, 2, ..., q\}$, $\tilde{\varepsilon} = \{\varepsilon_1, ..., \varepsilon_N\}$ and $\tilde{I} = \{I_i, i = 1, ..., N\}$. $\tilde{\varepsilon}$ is independent of $\tilde{I}$ given $\mathbf{Z}$, thus $E(2I_i - 1)\varepsilon_i = 0$ and

$$\frac{2}{N}E(\sum_i (2I_i - 1)\varepsilon_i | \mathbf{Z}) = 0,$$

and

$$Var(\frac{2}{N}(\sum_i (2I_i - 1)\varepsilon_i | \mathbf{Z}) = \frac{4\sigma_\varepsilon^2}{N}.$$

By the central limit theorem, given $(\tilde{I}, \mathbf{Z})$,

$$\frac{2}{\sqrt{N}}(\sum_i (2I_i - 1)\varepsilon_i | \mathbf{Z}) \xrightarrow{P} N(0, 4\sigma_\varepsilon^2)$$

and,

$$\frac{2}{\sqrt{N}} \sum_j \sum_i (2I_i - 1)\beta_j Z_{i,j} \xrightarrow{P} 0.$$

By Slutsky theorem,

$$\sqrt{N}A \xrightarrow{D} N(0, 4\sigma_\varepsilon^2).$$

Next, we will show $\sqrt{N}B \xrightarrow{P} 0$,

$$\sqrt{N}B = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})(\frac{\mathbf{X^T Z}}{\sqrt{N}}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{\sqrt{N}}).$$

Notice that

$$\frac{\mathbf{X^T Z}}{\sqrt{N}}\tilde{\beta} + \frac{\mathbf{X}^T\tilde{\varepsilon}}{\sqrt{N}} = \frac{1}{\sqrt{N}} \begin{bmatrix} \sum_j \sum_i I_i Z_{i,j}\beta_i + \sum_i I_i\varepsilon_i \\ \sum_j \sum_i (1 - I_i) Z_{i,j}\beta_i + \sum_i (1 - I_i)\varepsilon_i \\ \sum_j \sum_i X_{i,1} Z_{i,j}\beta_j + \sum_i X_{i,1}\varepsilon_i \\ \vdots \\ \sum_j \sum_i X_{i,p} Z_{i,j}\beta_j + \sum_i X_{i,p}\varepsilon_i \end{bmatrix}.$$

Then,

$$\sqrt{N}B = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})\frac{1}{\sqrt{N}}\begin{bmatrix} \sum_j \sum_i I_i Z_{i,j}\beta_i + \sum_i I_i\varepsilon_i \\ \sum_j \sum_i (1-I_i) Z_{i,j}\beta_i + \sum_i (1-I_i)\varepsilon_i \\ \sum_j \sum_i X_{i,1} Z_{i,j}\beta_j + \sum_i X_{i,1}\varepsilon_i \\ \vdots \\ \sum_j \sum_i X_{i,p} Z_{i,j}\beta_j + \sum_i X_{i,p}\varepsilon_i \end{bmatrix}.$$

$\sqrt{N}B$ can be written to two parts: S_1 and S_2, where

$$S_1 = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})\frac{1}{\sqrt{N}}\begin{bmatrix} \sum_j \sum_i I_i Z_{i,j}\beta_i + \sum_i I_i\varepsilon_i \\ \sum_j \sum_i (1-I_i) Z_{i,j}\beta_i + \sum_i (1-I_i)\varepsilon_i \\ \sum_i X_{i,1}\varepsilon_i \\ \vdots \\ \sum_i X_{i,p}\varepsilon_i \end{bmatrix}$$

and,

$$S_2 = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})\frac{1}{\sqrt{N}}\begin{bmatrix} 0 \\ 0 \\ \sum_j \sum_i X_{i,1} Z_{i,j}\beta_j \\ \vdots \\ \sum_j \sum_i X_{i,p} Z_{i,j}\beta_j \end{bmatrix}.$$

First, we will prove $S_1 = o_p(1)$,

$$\begin{aligned} \tfrac{1}{\sqrt{N}}(\textstyle\sum_j \sum_i I_i Z_{i,j}\beta_j + \sum_i I_i \varepsilon_i) &= \tfrac{1}{2}[\tfrac{1}{\sqrt{N}} \textstyle\sum_j \sum_i Z_{i,j}\beta_j + \tfrac{1}{\sqrt{N}} \sum_i \varepsilon_i \\ &\quad + \tfrac{1}{\sqrt{N}} \textstyle\sum_j \sum_i (2I_i - 1) Z_{i,j}\beta_j + \tfrac{1}{\sqrt{N}} \sum_i (2I_i - 1)\varepsilon_i]. \end{aligned}$$

By central limit theorem,

$$\frac{1}{\sqrt{N}} \sum_j \sum_i Z_{i,j}\beta_j + \frac{1}{\sqrt{N}} \sum_i \varepsilon_i = O_p(1).$$

Furthermore,

$$\frac{1}{\sqrt{N}} \sum_j \sum_i (2I_i - 1) Z_{i,j}\beta_j + \frac{1}{\sqrt{N}} \sum_i (2I_i - 1) = \frac{\sqrt{N}}{2} A.$$

Since $\frac{\sqrt{N}}{2}A$ converges to a normal distribution,

$$\frac{1}{\sqrt{N}}\sum_j\sum_i(2I_i-1)Z_{i,j}\beta_j+\frac{1}{\sqrt{N}}\sum_i(2I_i-1)\varepsilon_i=O_p(1).$$

Hence,

$$\frac{1}{\sqrt{N}}(\sum_j\sum_i I_iZ_{i,j}\beta_j+\sum_i I_i\varepsilon_i)=O_p(1).$$

Also, by symmetry,

$$\frac{1}{\sqrt{N}}(\sum_j\sum_i(1-I_i)Z_{i,j}\beta_j+\sum_i(1-I_i)\varepsilon_i)=O_p(1)$$

and since

$$(\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}-M^{-1}\xrightarrow{P}0.$$

Thus

$$S_1=L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1}-M^{-1})\frac{1}{\sqrt{N}}\begin{bmatrix}\sum_j\sum_i I_iZ_{i,j}\beta_i+\sum_i I_i\varepsilon_i\\ \sum_j\sum_i(1-I_i)Z_{i,j}\beta_i+\sum_i(1-I_i)\varepsilon_i\\ \sum_i X_{i,1}\varepsilon_i\\ \vdots\\ \sum_i X_{i,p}\varepsilon_i\end{bmatrix}=o_p(1).$$

Now, consider about S_2,

$$S_2 = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})\frac{1}{\sqrt{N}}\begin{bmatrix} 0 \\ 0 \\ \underline{X}Z \end{bmatrix}.$$

The formula for matrix inverse:

$$\begin{bmatrix} \mathcal{A} & \mathcal{U} \\ \mathcal{V} & \mathcal{C} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathcal{A}-\mathcal{U}\mathcal{C}^{-1}\mathcal{V})^{-1} & -(\mathcal{A}-\mathcal{U}\mathcal{C}^{-1}\mathcal{V})^{-1}\mathcal{U}\mathcal{C}^{-1} \\ -\mathcal{C}^{-1}\mathcal{V}(\mathcal{A}-\mathcal{U}\mathcal{C}^{-1}\mathcal{V})^{-1} & \mathcal{C}^{-1}\mathcal{V}(\mathcal{A}-\mathcal{U}\mathcal{C}^{-1}\mathcal{V})^{-1}\mathcal{U}\mathcal{C}^{-1}+\mathcal{C}^{-1} \end{bmatrix}.$$

Let

$$\mathcal{A} = \begin{bmatrix} \sum I_i/N & 0 \\ 0 & \sum(1-I_i)/N \end{bmatrix}$$

$$\mathcal{U} = \mathcal{V}^T = \begin{bmatrix} \sum I_i X_{i,1}/N & \dots & \sum I_i X_{i,p}/N \\ \sum(1-I_i)X_{i,1}/N & \dots & \sum(1-I_i)X_{i,p}/N \end{bmatrix},$$

and

$$\mathcal{C} = \begin{bmatrix} \sum X_{i,1}X_{i,1}/N & \dots & \sum X_{i,1}X_{i,p}/N \\ \vdots & \ddots & \vdots \\ \sum X_{i,p}X_{i,1}/N & \dots & \sum X_{i,p}X_{i,p}/N \end{bmatrix}.$$

After matrix calculation,

$$S_2 = L((\frac{\mathbf{X}^T\mathbf{X}}{N})^{-1} - M^{-1})\frac{1}{\sqrt{N}}\begin{bmatrix} 0 \\ 0 \\ \\ \underline{X}Z \end{bmatrix} = (1,-1)[-(\mathcal{A}-\mathcal{U}\mathcal{C}^{-1}\mathcal{V})^{-1}\mathcal{U}\mathcal{C}^{-1}]\frac{1}{\sqrt{N}}\underline{X}Z.$$

It is easy to get

$$\mathcal{A}-\mathcal{U}\mathcal{C}^{-1}\mathcal{V} \xrightarrow{P} \begin{bmatrix} 1/2 & 0 \\ \\ 0 & 1/2 \end{bmatrix}.$$

Also by central limit theorem,

$$\sqrt{N}\mathcal{U} = \sqrt{N}\begin{bmatrix} \sum I_i X_{i,1}/N & \dots & \sum I_i X_{i,p}/N \\ \\ \sum(1-I_i)X_{i,1}/N & \dots & \sum(1-I_i)X_{i,p}/N \end{bmatrix} = O_p(1).$$

Thus

$$-(\mathcal{A}-\mathcal{U}\mathcal{C}^{-1}\mathcal{V})^{-1}\mathcal{U}\mathcal{C}^{-1} = \begin{bmatrix} 1/2 & 0 \\ \\ 0 & 1/2 \end{bmatrix}\mathcal{U}\Sigma_{x,x}^{-1} + o_p(N^{-1/2}).$$

In sum, we get

$$S_2 = \frac{2}{\sqrt{N}}(\sum(2I_i-1)X_{i,1}, \dots, \sum(2I_i-1)X_{i,p})\Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta} + o_p(N^{-1/2}).$$

By condition (2),

$$\sum(2I_i-1)X_{i,1} = o_p(N^{1/2}).$$

Thus

$$S_2 = o_p(1),$$

we finish the proof of

$$\sqrt{N}B \xrightarrow{P} 0.$$

It follows that

$$\sqrt{N}L(\hat{\alpha} - \tilde{\alpha}) = \sqrt{N}(A + B) = \frac{2}{\sqrt{N}}(\sum_i (2I_i - 1)\varepsilon_i | \mathbf{Z}) + o_p(1).$$

Therefore,

$$\sqrt{N}L(\hat{\alpha} - \tilde{\alpha}) \xrightarrow{D} N(0, 4\sigma_\varepsilon^2).$$

Now, consider about the denominator part, $\hat{\sigma}^2 L(\mathbf{X^T X})^{-1} L^T$. We will show it is larger than $4\sigma_\varepsilon^2$) which means it is an inflated estimator of the variance of $L\hat{\alpha}$.
It is easy to get,

$$L(\mathbf{X^T X})^{-1} L^T = \frac{4}{N} + o_p(\frac{1}{N})$$

and similar with chapter 2,

$$\begin{aligned} \hat{\sigma}^2 &= \tfrac{1}{N-p-2}(\tilde{Y} - \mathbf{X}\hat{\alpha})^T(\tilde{Y} - \mathbf{X}\hat{\alpha}) \\ &= \tfrac{1}{N}[(\tilde{Y} - \mathbf{X}\tilde{\alpha})^T(\tilde{Y} - \mathbf{X}\tilde{\alpha}) + (\hat{\alpha} - \tilde{\alpha})^T\mathbf{X}^T\mathbf{X}(\hat{\alpha} - \tilde{\alpha}) - (\hat{\alpha} - \tilde{\alpha})\mathbf{X}^T(\tilde{Y} - \mathbf{X}\tilde{\alpha})] + o_p(1). \end{aligned}$$

From Theorem 3.3.1,

$$\frac{1}{N}(\hat{\alpha}-\tilde{\alpha})^T\mathbf{X}^T\mathbf{X}(\hat{\alpha}-\tilde{\alpha}) \xrightarrow{P} (0,0,\tilde{\beta}^T\Sigma_{x,z}^T\Sigma_{x,x}^{-1})*diag(\frac{1}{2},\frac{1}{2},\Sigma_{x,x}) \begin{bmatrix} 0 \\ 0 \\ \Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta} \end{bmatrix} = \tilde{\beta}^T\Sigma_{x,z}^T\Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta},$$

moreover,

$$\frac{2}{N}(\hat{\alpha}-\tilde{\alpha})^T\mathbf{X}^T(\tilde{Y}-\mathbf{X}\tilde{\alpha}) = 2(\hat{\alpha}-\tilde{\alpha})^T\frac{\mathbf{X}^TZ\tilde{\beta}}{N} + 2(\hat{\alpha}-\tilde{\alpha})^T\frac{\mathbf{X}^T\tilde{\varepsilon}}{N} \xrightarrow{P} 2\tilde{\beta}^T\Sigma_{x,z}^T\Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta},$$

furthermore,

$$\frac{1}{N}(\tilde{Y}-\mathbf{X}\tilde{\alpha})^T(\tilde{Y}-\mathbf{X}\tilde{\alpha}) = \frac{1}{N}(\mathbf{Z}\tilde{\beta}+\tilde{\varepsilon})^T(\mathbf{Z}\tilde{\beta}+\tilde{\varepsilon}) \xrightarrow{P} \sigma_\varepsilon^2 + \tilde{\beta}^T\Sigma_{z,z}\tilde{\beta}.$$

Thus, in sum

$$\hat{\sigma}^2 \xrightarrow{P} \sigma_\varepsilon^2 + \tilde{\beta}^T\Sigma_{z,z}\tilde{\beta} - \tilde{\beta}^T\Sigma_{x,z}^T\Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta},$$

and

$$\hat{\sigma}^2L(\mathbf{X^TX})^{-1}L^T = \frac{4}{N}(\sigma_\varepsilon^2 + \tilde{\beta}^T\Sigma_{z,z}\tilde{\beta} - \tilde{\beta}^T\Sigma_{x,z}^T\Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta}) + o_p(\frac{1}{N}).$$

Also,

$$\hat{\sigma}^2L(\mathbf{X^TX})^{-1}L^T = \frac{4}{N}(\sigma_\varepsilon^2 + Var(\mathbf{Z}^T - \mathbf{X}^T\Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta})) + o_p(\frac{1}{N}).$$

Obviously,

$$\sigma_\varepsilon^2 + Var(\mathbf{Z}^T - \mathbf{X}^T\Sigma_{x,x}^{-1}\Sigma_{x,z}\tilde{\beta}) > \sigma_\varepsilon^2,$$

which means $\hat{\sigma}^2L(\mathbf{X^TX})^{-1}L^T$ is inflated variance.

Under $H_0 : \mu_1 - \mu_2 = 0$,

$$T \xrightarrow{D} N(0, \tau^2), \tau^2 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + Var(\mathbf{Z}^T - \mathbf{X}^T \Sigma_{x,x}^{-1} \Sigma_{x,z} \tilde{\beta})},$$

for any $\beta_j \neq 0, \sigma_\varepsilon^2 < \sigma_\varepsilon^2 + Var(\mathbf{Z}^T - \mathbf{X}^T \Sigma_{x,x}^{-1} \Sigma_{x,z} \tilde{\beta})$.

And when $N \to \infty$,

$$Pr(|T| > Z_{(1-\alpha/2)}) \to 2\Phi(-\frac{\sigma_z Z_{1-\alpha/2}}{\sigma_\varepsilon}) < \alpha.$$

Under $H_a : \mu_1 - \mu_2 \neq 0$, $\mu_1 - \mu_2 = \delta/\sqrt{N}$ for fixed $\delta \neq 0$,

$$T \xrightarrow{D} N(\Delta, \tau^2), \ where \ \Delta = \frac{\delta}{2\sigma_z}.$$

Finish the proof of Theorem 3.1.2

□

Chapter 4

Proposed covariate-adaptive randomization with continuous covariate

Different methods have been published to balance continuous covariates. Most of these balancing procedures are designed to balance continuous covariates by minimizing the difference in certain characteristics of the continuous covariates among treatments. However, characteristics usually cannot capture the whole distribution. In real clinical trials, the distributions of the fatal baseline factors are usually not those common distributions, like normal or exponential distribution.

Kernel density estimator is an accepted and well-established nonparametric estimator for continuous covariate density. Ma and Hu (2013) proposed a method to balance continuous covariates sequentially using a biased coin in favor of the treatment with lower covariate density at the local covariate value of the incoming patient. However, the method proposed by Ma and Hu is not a minimization procedure, it does not consider the imbalance measure after assigning the incoming

patient. Also based on the kernel density estimator of the continuous covariate densities, we proposed a modified kernel density randomization which is a minimization method.

Consider a clinical trial with two treatments: 1 and 2, the patients are enrolled sequentially. Suppose for each patient, M covariates need to be balanced, among which m_1 are continuous and m_2 are categorical. Denote $Z_i = (Z_{i1}, \cdots, Z_{iM})$, as the covariate profile of the i-th patient. Assume the first m_1 covariates $(Z_{i1}, \cdots, Z_{im_1})$ are continuous, and the last m_2 covariates $(Z_{i(m_1+1)}, \cdots, Z_{iM})$ are categorical. Suppose after assigning treatment to the first n patients sequentially, n_1 patients have been allocated to treatment 1 and n_2 to treatment 2, where $n_1 + n_2 = n$. Moreover, let $Z_i^{(k)} = (Z_{i1}^{(k)}, \cdots, Z_{iM}^{(k)})$, $i = 1, \cdots, n_k$, denote the covariate of patients in treatment k, $k = 1, 2$. When patient $n+1$ enters the trial with covariate profile Z_{n+1}, we determine the allocation of the new patient to achieve better similarity of covariate distribution among the two treatments.

4.1 Minimization procedure

4.1.1 Kernel density estimate

A novel covariate-adaptive allocation procedure is proposed to minimize distributions for both continuous and discrete covariates between two treatments. For a continuous covariate Z_j, $j = 1, \cdots, m_1$, its density function in treatment k, $k = 1, 2$, can be estimated as

$$\hat{f}_{j,k}(z) = \frac{1}{n_k h(n_k)} \sum_{i=1}^{n_k} K(\frac{z - Z_{ij}^{(k)}}{h(n_k)})$$

where $K(\cdot)$ is the kernel. In our method, normal kernel is used,

$$K(x) = \frac{1}{\sqrt{2\pi}} exp(-x^2).$$

The bandwidth $h(n_k)$ is a function of the sample size n_k. Here, we use Scott's rule: $h(n_k) = \hat{\sigma} n_k^{-0.2}$, where $\hat{\sigma}$ is the estimated standard deviation of the covariate. The mean integrated squared error can be minimized when use Scott's rule.

As the original kernel method proposed, the density estimation is proportion-adjusted to maintain balance measure in group size.

$$d_{j,k}(z) = \frac{n_k}{n} \hat{f}_{j,k}(z).$$

Note that the kernel density estimation for covariate j based on all patients in two treatments is

$$\hat{f}_j(z) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{z - Z_{ij}}{h_n}).$$

By choosing the same bandwidth like $h(n_k) = h_n$, $k = 1, 2$, we have the following density decomposition of covariate j

$$\hat{f}_j(z) = d_{j,1}(z) + d_{j,2}(z).$$

Therefore, $d_{j,k}(z)$ is actually the proportional density of covariate j that decomposed to treatment k.

For discrete covariate j, $j = m_1 + 1, \cdots, M$, distribution is captured by the relative frequency. Suppose covariate j has E_j levels and patient $n + 1$ falls in the level $Z_{(n+1)j}$, $1 \leq Z_{(n+1)j} \leq E_j$. Within level $Z_{(n+1)j}$ of covariate Z_j, assume

n_{j1} patients have been assigned to treatment 1 and n_{j2} assigned to treatment 2, respectively. The density of covariate j for treatment k on the level that the new patient falls is, $k = 1, 2$,

$$d_{j,k}(Z_{(n+1)j}) = n_{jk}/n.$$

Similar with continuous covariates, $d_{j,k}$ is the density with respect to total patients in two treatments.

4.1.2 Imbalance measure

When a new patient is ready to join a clinical trial with two treatment: 1 and 2, we calculate the kernel density estimators by potentially assigning a new patient to a treatment group. Let $Z_{(n+1)j}$ the jth covariate of the new patient with value z_0, thus $Z_{(n+1)j} = z_0$.

If the new patient is assigned to treatment 1, and the density function is defined as

$$\hat{f}^{(1)}_{j,1}(z_0) = \frac{1}{(n_1+1)h(n_1+1)} \sum_{i=1}^{n_1+1} K(\frac{z_0 - z_{i,j}}{h(n_1+1)})$$

$$\hat{f}^{(1)}_{j,2}(z_0) = \frac{1}{(n_2)h(n_2)} \sum_{i=1}^{n_2} K(\frac{z_0 - z_{i,j}}{h(n_2)}),$$

then the proportion-adjusted kernel density estimator will be:

$$d^{(1)}_{j,1}(z_0) = \frac{n_1+1}{n+1}\hat{f}^{(1)}_{j,1}(z) \quad and \quad d^{(1)}_{j,2}(z_0) = \frac{n_2}{n+1}\hat{f}^{(1)}_{j,2}(z).$$

The imbalance measure $\Delta d^{(1)}_j$ is defined as the following equation if patient $n+1$ is

assigned to treatment 1,

$$\Delta d_j^{(1)} = |d_{j,1}^{(1)}(z_0) - d_{j,2}^{(1)}(z_0)|$$

as the density difference with regard to covariate j for the patient $n+1$. Similarly, if patient $n+1$ is assigned to treatment 2, then the imbalance measure $\Delta d_j^{(2)}$ is denoted as:

$$\Delta d_j^{(2)} = |d_{j,1}^{(2)}(z_0) - d_{j,2}^{(2)}(z_0)|$$

as the density difference with in regard to the discrete covariate j for the new patient.

For discrete covariate Z_j, $j = m_1+1, \cdots, M$, assume patient $n+1$ with covariate Z_j falls in the level z_0, say the gender: male or female. Within level z_0 of covariate Z_j, let n_{j1} and n_{j2} be the number of patients that have been allocated to treatment 1 and treatment 2 respectively. If the new patient is assigned to treatment 1, then the density of covariate j for treatment 1 will be:

$$d_{j,1}^{(1)}(z_0) = (n_{j1}+1)/(n+1).$$

The density for treatment 2 will be:

$$d_{j,2}^{(1)}(z_0) = n_{j2}/(n+1).$$

Similar with continuous covariates, the imbalance measure for discrete covariate j by assigning new patient to treatment 1 is:

$$\begin{aligned} \Delta d_j^{(1)} &= |d_{j,1}^{(1)}(z_0) - d_{j,2}^{(1)}(z_0)| \\ &= |(n_{j1}+1-n_{j2})/(n+1)|, \end{aligned}$$

where $\Delta d_j^{(1)}$ is the density difference in regard to the discrete covariate j for the new patient. The imbalance measure for discrete covariate j by assigning new patient to treatment 2:

$$\begin{aligned} \Delta d_j^{(2)} &= |d_{j,1}^{(2)}(z_0) - d_{j,2}^{(2)}(z_0)| \\ &= |(n_{j1} - n_{j2} - 1)/(n+1)|, \end{aligned}$$

where $\Delta d_j^{(2)}$ is the density difference in regard to the discrete covariate j for the new patient.

Finally, we define the imbalance measure Imb as the weighted average of Δd_j, $j = 1, \cdots, M$.

$$Imb = \sum_{j=1}^{M} w_j (\Delta d_j^{(1)} - \Delta d_j^{(2)})$$

where w_j is a nonnegative weight placed on covariate j. with nonnegative weights w_j such that

$$\sum_{j=1}^{M} w_j = 1.$$

In practice, the variance of covariates may differ from each other, which means the weight cannot reflect the relative importance of each covariate. It is recommended that each continuous covariate should be standardized with the same variance before they are randomized. Then we can put heavier weight to the covariates with comparative importance. In real clinical trials, the variance of covariates could be estimated and adjusted sequentially.

When patient $n+1$ enters the trial with covariate Z_{n+1}, we determine the allocation of this new patient using a biased coin, which favors the treatment with lower density at Z_{n+1}. Let p be a biased probability, such that $0.5 < p < 1$, the procedure of biased coin allocation for continuous covariate can be summarized as follow,

Step 1 : Initial step

Assign n_0 patients to each treatment by using a restricted randomization.

Step 2 : Imbalance calculation

Suppose $n(n > 2n_0)$ patients have been assigned to treatments, calculate the imbalance measure Imb.

Step 3 : Biased allocation

If $Imb < 0$, assign patient $n+1$ to treatment 1 with probability p.

If $Imb > 0$, assign patient $n+1$ to treatment 2 with probability p.

If $Imb = 0$, assign patient $n+1$ to treatment 1 or 2 with probability $1/2$.

The proposed procedure reduces to Pocock and Simon's minimization when all covariates are discrete.

4.2 Simulation study

In our simulation studies the adjusted kernel density method is compared with other continuous covariate minimization procedures from the literature. Different scenarios regarding the number and nature of covariates are considered. In the first study, the simplest case of a single normally distributed covariate is considered. The new randomization method is compared with other procedures in various perspectives. In our second simulation study, two and more covariates are incorporated in the randomization procedure. Besides the popular covariate-adaptive randomization methods, the two most widely used methods: the stratified permuted block design (SPBD) and the Pocock and Simon's method, are also considered here by discretizing continuous covariates into categories.

We evaluate the balance of a clinical trial from various perspectives: balance of group size, mean, median and balance of covariate distribution. For discrete covariate, the p-value of χ^2 test is used to measure the degree of balance. The larger the p-value is, the more balance of a covariate among two treatments. For continuous covariate, Kolmogorov-Smirnov (K-S) test is commonly used to test the equality of distribution. Thus the p-value of Kolmogorov-Smirnov (K-S) test is used to measure the degree of balance of continuous covariates.

4.2.1 Balancing single continuous covariate

Our new kernel method is compared with the kernel density method (Kernel) proposed by Ma and Hu, the Stratified permuted block design (SPBD), the Pocock and Simon's (P-S) method, the Rank-sum test method (Rank), p-value method (P-value) by Frane, the Kullback-Leibler divergence measure (KLD) by Endo, quartiles minimization (Quartile) by Su, empirical cumulative distribution (ECDF) by Lin and Su, and the Complete Randomization (CR). (The details of above covariate-adaptive designs can be found in section 1.2.2)

Study 1 considers an idealized clinical trial with one continuous covariate from standard normal distribution for 100 subjects. Simulation results were based on 5,000 repetitions. The biased coin allocation with $p = 0.8$ was used for all methods. Median case 50 percentile of overall difference $|n_1 - n_2|$, K-S test, Δmean, Δmedian, as well the worst case 99 percentile are checked.

For Pocock and Simon's method and stratified permuted block design, the normal covariate is divided into three categories by cutting on 1 and -1. The squared difference in numbers is used in P-S method. For P-value, KLD, ECDF, Quantile methods, 6 is setted to be the maximum tolerant imbalance of $|n_1 - n_2|$.

Table 4.1: Comparison of overall imbalance and p-value of K-S test for methods: New Kernel density method (New Kernel), Pocock and Simon's marginal procedure (P-S), Stratified permuted block design (SPBD), Rank-sum method (Rank), p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR), simulation based on 5,000 runs.

	50 Percentile		99 (1) Percentile	
Method	ΔN	K-S	ΔN	K-S
New Kernel	2	0.972	6	0.587
P-S	0	0.720	4	0.108
SPBD	0	0.717	4	0.048
P-value	4	0.743	6	0.112
KLD	4	0.805	6	0.158
Rank	4	0.548	12	0.022
Quartile	2	0.869	6	0.272
ECDF	4	0.986	6	0.641
Kernel	2	0.975	6	0.563
CR	6	0.540	24	0.007

In controlling total patient number difference between two groups, Pocock and Simon's method and stratified permuted block design have the best performance. When the number of strata is small, stratified permuted block design gives good within stratum balance and overall balance. The p-value of K-S test shows, ECDF method and two Kernel methods achieve the minimum distribution difference among two treatments. Pocock and Simon's method and stratified permuted block design do not perform as well as covariate-adaptive designs here. Complete randomization works worst in both overall patient number difference and distribution difference.

Table 4.2: Comparison of the difference in mean and median for methods: New Kernel density method (New Kernel), Pocock and Simon's marginal procedure (P-S), Stratified permuted block design (SPBD), Rank-sum method (Rank), p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR), simulation based on 5,000 runs.

	50 Percentile		99 Percentile	
Method	Δmean	Δmedian	Δmean	Δmedian
New Kernel	0.055	0.058	0.206	0.252
P-S	0.083	0.058	0.315	0.295
SPBD	0.078	0.133	0.283	0.534
P-value	0.016	0.114	0.093	0.421
KLD	0.024	0.105	0.128	0.403
Rank	0.123	0.155	0.460	0.602
Quartile	0.064	0.080	0.238	0.326
ECDF	0.031	0.047	0.134	0.220
Kernel	0.058	0.062	0.225	0.245
CR	0.119	0.151	0.525	0.629

In perspective of mean difference, P-value method and KLD method achieve the minimum imbalance. Both kernel method and our new kernel method also perform well in controlling mean difference. The reason is that the imbalance measure for P-value method is the ratio of mean difference and weighted variance. KLD method targets on minimizing mean difference and variance difference for normal covariate. In controlling median difference, ECDF method and two Kernel methods are the best. Because these methods target on balancing distributions of continuous covariates among groups, thus have good balance in respect to mean and median.

4.2.2 Balancing multiple covariates

Study 2: consider a clinical trial with two treatments, apply various covariate-adaptive methods to balance 4 baseline factors for 200 subjects. Assume two continuous covariates are from normal and exponential, i.e., $Z_1 \sim N(0,1)$, $Z_2 \sim Exp(1)$, representing symmetric and skewed distributions respectively. Two discrete covariates are from Bernoulli distribution with $p = 0.5$, i.e., $Z_3,\ Z_4 \sim Binary(0.5)$. For Pocock and Simon's (P-S) method and stratified permuted block design (SPBD), the normal covariate was divided into three categories by cutting on 1 and -1, while exponential covariate is stratified at mean 1. The square of patient number difference among groups is used in P-S method. For P-value, KLD, ECDF, Quartile methods 6 is setted to be the maximum tolerant imbalance of group size.

Simulation results were based on 5,000 repetitions. The biased coin allocation with $p = 0.8$ was used for all methods. Median case 50-percentile of overall difference $|n_1 - n_2|$, K-S test for continuous covariates, χ^2 test for discrete covariates, as well the worst case 99 percentile are checked.

Table 4.3: Comparison of overall imbalance and p-value of K-S test for methods: Pocock and Simon's marginal procedure (P-S), Stratified permuted block design (SPBD), Rank-sum method (Rank), p-value based randomization (P-value), Su's percentile method (Quartile), Empirical cumulative distribution function method (ECDF), Kullback-Leibler divergence method (KLD), Kernel density procedure (Kernel) and Complete Randomization (CR) for 4 covariate, 200 subjects, simulation based on 5,000 runs.

	50 Percentile				99 Percentile			
Method	ΔN	Nor	Exp	Bin	ΔN	Nor	Exp	Bin
New Kernel	2	0.902	0.906	0.908	4	0.317	0.296	0.654
P-S	0	0.729	0.710	0.894	4	0.075	0.066	0.639
SPBD	4	0.692	0.699	0.838	10	0.043	0.045	0.394
KLD	4	0.812	0.721	0.887	6	0.199	0.102	0.494
Rank	2	0.691	0.699	0.677	10	0.083	0.088	0.109
Quartile	0	0.813	0.775	0.790	4	0.124	0.078	0.152
ECDF	4	0.838	0.812	0.907	6	0.197	0.102	0.677
Kernel	2	0.912	0.906	0.904	4	0.281	0.275	0.683
CR	6	0.518	0.518	0.617	26	0.012	0.012	0.020

Table 4.3 summarized the overall group size difference in two treatments, p-value of K-S test for normal covariate, p-value of K-S test for exponential covariate and p-value of χ^2 test for categorical covariates. It shows, kernel density method and our new kernel method can achieve largest p-value for both continuous covariates as well as good balance in discrete covariates.

Chapter 5

Conclusions and future work

5.1 Conclusion

In a covariate-adaptive design, the covariates of patients in groups are balanced and kept randomized, which is important when comparing treatment effectiveness in clinical trials. While most clinical trials still use permuted-block design, minimization methods are becoming more widely accepted. Current minimization techniques still discretize continuous covariates for randomization at the cost of losing information. More recent works in the literature have proposed various continuous covariate-adaptive designs targeting balance specific characteristics in the distribution of the covariates for different groups. Ma and Hu (2013) proposed a design based on the kernel-density estimate. Their method out-performs a variety of existing continuous-covariate minimization methods for randomization including the p-value based method, KLD method, quantile method and ECDF method.

However, Ma and Hu's proposed kernel-density method is not a minimization method for randomizing patient groups. Our proposed work is minimization adaptation of the kernel-density method. We show our adaptation performs just as well

as the Ma and Hu's method and better than various existing covariate-adaptive designs. The advantage of the kernel-density method compared to other widely used covariate adaptive randomization methods is the statistical power in minimizing distribution-differences of covariates between two treatments. The kernel-density method ensures proper balance of covariates among groups with mean, median and overall distributions. While other continuous covariate-randomization methods require thresholds on group sample size, the kernel-density method ensures two group sample size difference is bounded in probability even without using it as a part of the imbalance measure to do randomization.

In chapter two, we discuss hypothesis testing for clinical trials based on various continuous covariate randomizations. Assuming a linear relationship for the response and covariates, we conduct hypothesis testing to determine the difference between treatment effects and the significance of covariates. When the conditions in theorem 2.2.1 hold, the hypothesis testing of treatment effect under null hypothesis is conservative. However if the trial uses complete randomization to allocate patients then the hypothesis testing of treatment effect difference is valid. The power of hypothesis testing is smaller for covariate-adaptive design than complete randomization when the difference is small and sample size is small. As the difference between null hypothesis and underlying truth increases and sample size increases, the power will also increase. Moreover, the power of covariate-adaptive design is larger than complete randomization. Meanwhile hypothesis testing for covariate significance is always valid regardless of the randomization method used in the clinical trial.

5.2 Future work

For future works the properties of the hypothesis testing can be generalized to many other ways under covariate-adaptive design. Beyond the current linear model, we can adapt to more complex situations using a more general model, e.g., logistic model. While our proposed method handles only two treatments, generalizing to handle more than two methods is important. Finally, we assume the covariates are independent, but in real clinical trials covariates may be correlated; therefore, incorporating their correlation is an important next step to the current proposed method.

In hypothesis testing problems under covariate-adaptive designs, p-value based method and KLD method give the most conservative testing results among all the continuous covariate-adaptive designs. In particular, the test statistic for the p-value based method minimizes the ratio of mean difference and weighted variance, while the KLD method minimizes the mean difference and variance difference in cases of normal covariates. For example, minimizing an imbalance measure that contains the mean difference results in the least difference between group means. Both p-value based method and KLD method with smaller group mean difference lead to smaller type-I errors of the hypothesis testing. Simulations shows $\sum_{i=1}^{n_1} Z_{i,1} - \sum_{i=1}^{n_2} Z_{i,2} = O_p(1)$ in both p-value based method and KLD method. Moreover, all the continuous covariate-adaptive deigns discussed here should satisfy the condition 3 in theorem 2.2.1 which is $\sum_{i=1}^{n_1} Z_{i,1} - \sum_{i=1}^{n_2} Z_{i,2} = o_p(\sqrt{n})$.

For the continuous covariate-adaptive designs, the proposed approach of balancing continuous covariate shows desired properties in simulation. Investigation on the theoretical properties of the proposed approach would be interesting and important. Though various approaches to balance continuous covariate have been proposed in

the literature, none of them proved any theoretical properties.

To start with, let's consider the covariate-adaptive randomization with one continuous covariate. Let T_i denotes treatment assignment of patient i, with $T_i = 1$ if assigned to treatment 1 and $T_i = 0$ otherwise. After the randomization of the first n patients, we have the following density estimation for covariate Z in treatment 1 and 2. Here the density estimation has been adjusted for group size

$$d_1(z) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{z - Z_i}{h_n}) T_i$$

and

$$d_2(z) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{z - Z_i}{h_n})(1 - T_i).$$

Then the density difference between two groups can be measured as:

$$\Delta d(z) = d_1(z) - d_2(z) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{z - Z_i}{h_n})(2T_i - 1).$$

Simulations are carried out to investigate the convergence speed of $\Delta d(z)$. In particular, we consider the case that covariate Z follows standard normal distribution and run simulations based on different sample sizes. Density difference at different values of z are considered. Based on the simulation result, it's reasonable to propose the following hypothesis:

$$\Delta N \sim O_p(1) \quad and \quad (nh_n)\Delta d(z) \sim O_p(1).$$

For future works we hope to give a rigorous derivation for the above two theoretical properties of kernel density method.

Bibliography

[1] Aickin, M. (2009). A simulation study of the validity and efficiency of design-adaptive allocation to two groups in the regression situation. *The International Journal of Biostatistics*, **5**, 1-14 .

[2] Anderson, H. Hopwood, P., Stephens, R. J.,Thatcher, N., Cottier, B., Nicholson, M., Milroy, R., Maughan, T. S., Falk, S. J., Bond, M. G., Burt, P. A., Connolly, C. K., McIllmurray, M. B. , Carmichael, J. (2009). Gemcitabine plus best supportive care (BSC) vs BSC in inoperable non-small cell lung cancer-a randomized trial with quality of life as the primary outcome. *British Journal of Cancer*, **83**, 4, 447-453.

[3] Atkinson, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognoctic factors. *Biometrika*, **69**, 61-67.

[4] Baldi Antognini, A. and Giovagnoli, A. (2004). A new sbiased coin designs for the sequential allocation of two treatments. *Applied Statistics*, **53**, 651s-664.

[5] Begg, C. B. and Iglewicz, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrika*, **36**, 81-90.

[6] Biswas, A., Angers, J.-F. (2002). A Bayesian adaptive design in clinical trials for continuous responses. *Statistica Neerlandica*, **56**, 400?414.

[7] Blackwell D. and Hodges, J. L. (1957). Design for the control of selection bias. *Annals of Mathematical Statistics*, **28**, 449-460.

[8] Brikett, N. J. (1985). Adaptive allocation in randomized controlled trials. *Controlled Clinical Trials*, **6**, 146-155.

[9] Chen, Y. -P. (1999). Biased coin design with imbalance tolerance. *Communications in Statistics: Stochastic Models*, **15**, 953-975.

[10] Cheung, Y. K., Lurdes, Y. T., Wathen, J. K., Thall, P. F. (2006). Continuous Bayesian adaptive randomization based on event times with covariates. *Statistics in Medicine*, **25**, 55?70.

[11] Ciolino J., Zhao W., Martin R., Palesch Y. (2011). Quantifying the cost in power of ignoring continuous covariate imbalances in clinical trial randomization. *Contemporary Clinical Trials.*, **32**:250-259.

[12] Eisele, J. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference,* **38**, 249-262.

[13] Eisele, J. and Woodroofe, M. (1995). Central limit theorems for doubly adaptive biased coin designs. *The Annals of Statistics*, **23**, 234-254.

[14] Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika*, **58**, 403-417.

[15] Endo A., Nagatani F., Hanada C., Yoshimura A. (2006). Minimization method for balancing continuous prognostic variables between treatment and control groups using Kullback-Leibler divergence. *Contemporary Clinical Trials*, **27**:420-431.

[16] Forsythe, A. B. (1987). Validity and power of tests when groups have been balanced for prognostic factors. *Computational Statistics Data Analysis*, **5**, 193-200.

[17] Frane J.W. (1998). A method of biased coin randomization, its implementation, and its validation. *Drug Information Journal*, **32**:423-432.

[18] Green, S. B. and Byar, D. P. (1978). The effect of stratified randomization on size and power of statistical tests in clinical trials. *Journal of Chronic Diseases*, **31**, 445-454

[19] Hagino, A., Hamada, C., Yoshimura, I., Ohashi, Y., Sakamoto, J. and Nakazato, H. (2004). Statistical comparison of random allocation methods in cancer clinical trials. *Controlled Clinical Trials*, **25**, 572-584.

[20] Heritier, S., Gebski, V. and Pillai, A. (2005). Dynamic balancing randomization in controlled clinical trials. *Statistics in Medicine*, **24**, 3729-3741.

[21] Hoehler, F. (1987). Balancing allocation of subjects in biomedical research: a minimization strategy based on ranks. *Computers And Biomedical Research*, **20**, 209-213.

[22] Hu, F. (2012). Statistical issues in trial design and personalized medicine. *Clinical Investigation*, **2**, 121-124.

[23] Hu, Y. and Hu, F. (2012). Asymptotic properties of covariate-adaptive randomization. *Annals of Statistics*, **40**, 1794-1815.

[24] Hu, F. and Rosenberger, W. F. (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association*, **98**, 671-678.

[25] Hu, F., and Rosenberger, W. F. (2006). The theory of response-adaptive randomization in clinical trials. Vol. 525. John Wiley & Sons.

[26] Hu, F. and Zhang, L. (2013). On the theory of covariate-adaptive designs. *Annals of Statistics*, submitted.

[27] Hu, F., Zhang, L.-X., and He, X. (2009). Efficient randomized-adaptive designs. *Annals of Statistics* , **37**, 2543-2560.

[28] Ivanova, A. (2003). A play-the-winner type urn model with reduced variability. *Metrika*, **58**, 1-13.

[29] Ivanova, A., and N. Flournoy (2001). A birth and death urn for ternary outcomes: stochastic processes applied to urn models. Probability and Statistical Models with Applications: 583-600.

[30] Kalish, L. A. and Begg, C. B. (1985). Treatment allocation methods in clinical trials: a review. *Statistics In Medicine*, **4**, 129-144.

[31] Lin, Y. and Su, Z. (2012). Balancing continuous and categorical baseline covariates in sequential clinical trials using the area between empirical cumulative distribution functions. *Statistics in Medicine.*, **31**, 1961-1971.

[32] Nordle, O. and Brantmark, B. (1977). A self-adjusting randomization plan for allocation of patients into two treatment groups.*Clin Pharmacol Ther.*, **22(6)**, 825-830.

[33] Ma, W., Hu, F., and Zhang, L. (2014). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association*, **110**, 669-680.

[34] Ma, Z. and Hu, F. (2013). Balancing continuous covariates based on kernel densities. *Contemporary Clinical Trials.*, **34**, 262-269.

[35] McEntegart D. J. (2003). The pursuit of balance using stratified and dynamic randomization techniques: an overview *Drug Information Journal*, **37**, 2, 293-308.

[36] Molander, A., Warfvinge, J., Reit, C. and Kvist, T.(2012).Clinical and radiographic evaluation of one- and two-visit endodontic treatment of asymptomatic necrotic teeth with apical periodontitis: a randomized clinical trial. *Journal of Endodontics*, **33**, 10, 1145-1148.

[37] Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, **3**, 103-115

[38] Rosenberger and Lachin. (2002). Restricted Randomization. *Comput Biomed Res* , **15(6)**, 519-524.

[39] Rosenberger, W. F. and Sverdlov, O. (2008). Handling covariates in the design of clinical trials. *Statistical Science*, **23**, 404-419.

[40] Shao, J., Yu, X. and Zhong, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika*, **97**, 2, 347-360.

[41] Smith R.L., (1984). Properties of biased coin designs in sequential clinical trials. *Ann. Statist.*, **12**:1018-1034.

[42] Smith R.L., (1984). Sequential treatment allocation using biasing coin designs. *Journal of the Royal Statistical Society Series B*, **46**:519-543.

[43] Simon, R (1979). Restricted randomization designs in clinical trials. *Biometrics*, **35**, 2, 503-512.

[44] Signorini, D. F., Leung, O. , Simes, R. J., Beller, E., Gebski, V. J. and Callaghan, T. (1993). Dynamic balanced randomization for clinical trials. *Statistics in medicine*, **12**, 2343-2350.

[45] Soares, J. F. and Wu, C. J. F. (1983). Some restricted randomization rules in sequential designs. *Communications in Statistics A: Theory and Methods*, **17**, 2017-2034.

[46] Stigsby, B. and Taves, D. R. (2010) Rank-Minimization for balanced assignment of subjects in clinical trials. *Contemporary Clinical Trials*, **31**, 147-150.

[47] Su Z. (2011). Balancing multiple baseline characteristics in randomized clinical trials. *Contemporary Clinical Trials*, **32**:547-550.

[48] Taves, D. R. (1974). Minimization: a new method of assigning patients to treatment and control groups.*Clin Pharmacol Therap*, **15**, 443-453.

[49] Taves, D. R. (2010).The use of minimization in clinical trials *Contemporary Clinical Trials*, **31**, 180-184.

[50] Thall, P. F., Wathen, J. K. (2005). Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. *Statistics in Medicine*, **24**,1947?1964.

[51] Tymofyeyev, Y., Rosenberger, W.F. and Hu, F. (2007).Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*, **102**, 224-234.

[52] Tu, D., Shalay, K and Pater, J. (2000).Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. *Drug Information Journal,* **31**, 180-184.

[53] Wei, L. J. (1977). A class of designs for sequential clinical trials. *Journal of the American Statistical Association,* **72**, 382-386.

[54] Wei, L. J. (1978a). The adaptive biased coin design for sequential experiments. *Annals of Statistics*, **6**, 92-100

[55] Wei, L. J. (1978b). An application of an urn model to the design of sequential controlled clinical trials. *Journal of the American Statistical Association*, **73**, 559-563.

[56] Wei, L. J. and Durham, S. D. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, **73**, 840-843.

[57] Wei, L. J., Lachin, John M. (1988/12). Properties of the urn randomization in clinical trials. *Controlled Clinical Trials*, **9(4)**, 345-364.

[58] Weir, C. J. and Lees, K. R. (2003). Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Statistics In Medicine*, **22**, 705-726.

[59] Zelen, M. (1974).The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases*, **27**, 365-375.

[60] Zhang, L., Chan, W. S., Cheung, S. H. and Hu, F. (2007). A generalized urn model for clinical trials with delayed responses. *Statistica Sinica*, **17**, 387-409.

[61] Zhang, L., Hu, F. and Cheung, S. H. (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *Annals of Applied Probability*, **16**, 340-369.

[62] Zhang, L., Hu, F., Cheung, S, H. and Chan, W. S. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *The Annals of Statistics*, **35**, 1166-1182.

[63] Zhu, H., Hu, F. (2010). Sequential monitoring of response-adaptive randomized clinical trials. *Annals of Statistics*, **38**, 2218?2241.