This is to certify that the
dissertation entitled

RESAMPLING METHODS FOR ADAPTIVE DESIGNS

presented by

Hui Zhang

has been accepted towards fulfillment
of the requirements for the

Ph. D. degree in Statistics and Probability

Connie Page (CONNIE PAGE)

(Vincent Melfi)

Major Professor's Signature

12/7/05

Date

RESAMPLING METHODS FOR ADAPTIVE DESIGNS

By

Hui Zhang

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2005

ABSTRACT

RESAMPLING METHODS FOR ADAPTIVE DESIGNS

By

Hui Zhang

In clinical trials, the estimation of effects difference is often of primary importance. Proper resampling methods will provide second order correct estimates, which will outperform the traditional normal approximation. Bootstrapping has been known for a long time for i.i.d. random variables. Unfortunately, traditional bootstrapping methods are not appropriate because the observations of adaptive designs are dependent due to adaptive allocation. We address this problem by developing and studying resampling methods for dependent data in adaptive designs including theoretical results and simulations.

To my parents and Guoren

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis advisors Professor Vincent Melfi and Professor Connie Page for their invaluable assistance and insights leading to the writing of this dissertation. I am thankful for their constructive suggestions that significantly improved my unprofessional presentation of this work and brought it up to the present form. Their kindness and patience are very much appreciated. My sincere thanks also go to the members of my graduate committee, Professor L. Post and Professor R. Ramamoorthi, for their understanding and helpful suggestions.

Many thanks to Professor James Stapleton for always being there for me, providing help when I needed most and pulling me through my difficult times.

My research career is impossible without the patient love from my family. My parents Wenyan Zhang and Guoliang Zhang, my sister Rui Zhang have continually offered me support and encouragement.

Especially, I would like to give my special thanks to my husband Guoren Cheng, a genius in mathematics, for helping me trouble-shoot the simulations, stimulating suggestions, and turning my worries into solutions.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Introduction

In clinical trials, it is often desirable that design be adaptive using past information to allocate present subjects. For better statistical inference, resampling methods are often used to provide second order correct estimates. This dissertation is motivated by these two issues. The dissertation will focus on introducing and analyzing resampling methods for adaptive designs. Adaptive designs will be introduced in Section 1.1, and a short review of resampling methods for adaptive designs will be given in Section 1.2.

1.1 Introduction to Adaptive Designs

The main focus of this section is to introduce adaptive design and to review some of the adaptive designs that have already been proposed in the literature. Consider a clinical trial to evaluate the relative effectiveness of two treatments, A and B. It is assumed that patients arrive sequentially and each of the patients must be

assigned to exactly one of the two treatments. We assume that the response will be observed immediately. It may be desirable that the treatment assignment takes into consideration the information obtained from the past observations. A design that incorporates in the allocation rule the information obtained from past observations is called an *Adaptive Design* (For the benefit of readers, an index of definitions is given in Appendix A). A common use of adaptive designs is to compromise between two major yet conflicting goals: (i) to draw reliable statistical inference for the benefit of future subjects, which is the utilitarian goal; (ii) to maximize the total number of patients receiving the better treatment, which is the individualistic goal. Some early adaptive designs can be found in Thompson (1933) and Robbins (1952).

Adaptive designs can be divided into two groups: allocation adaptive and response adaptive designs.

- In allocation adaptive designs, the allocation rules are based only on the allocation of previous patients. For example, the *Biased Coin Design*, proposed by Efron (1971), is an allocation adaptive design.

- In response adaptive designs, the allocation rules are based on the responses as well as the allocations of previous patients. The allocation rules are either based on intuitive motivation, such as the *Randomized Play-the-winner Rule*, proposed by Wei and Durham (1978), or based on optimal target allocations, for instance, the *Doubly Adaptive Biased Coin Design*, proposed by Eisele(1994) and Eisele and Woodroofe (1995).

Applications of adaptive designs in many different disciplines are discussed in

Flournoy and Rosenberger (1995).

In the remainder of this dissertation, unless otherwise noted, we will consider an adaptive model in which two treatments are compared and the responses are binary. In Section 1.1.1, notation is introduced. Allocation adaptive designs are covered in Section 1.1.2, and response adaptive designs are covered in Section 1.1.3.

1.1.1 Notation

To better explain later work, it is necessary to introduce notation. In the setting of clinical trials, suppose there are two competing treatments, A and B. Each one of k sequentially arriving patients must be allocated to either treatment A or treatment B. Let X_j and Y_j represent the jth patient's immediate potential responses to treatments A and B, respectively, even though in practice only one of them will be observed. For simplicity, here and below, unless otherwise mentioned, we assume binary responses. Let p_A and p_B be the underlying probabilities of success of treatments A and B, respectively. It is assumed that the vectors $\{X_j, Y_j\}_{j=1}^{k}$ are i.i.d., where $X_1 \sim Bernoulli(p_A), Y_1 \sim Bernoulli(p_B)$. Note that there may be dependence within pairs. A sequential allocation procedure is given by a sequence of random variables $\{\delta_j\}_{j=1}^{k}$, where

$$\delta_j = \begin{cases} 1 & \text{if } X_j \text{ is observed;} \\ 0 & \text{if } Y_j \text{ is observed.} \end{cases} \tag{1.1}$$

The observation at stage j is given by $Z_j = \delta_j X_j + (1 - \delta_j) Y_j$. Let $\{U_j\}_{j=1}^{k}$ be a sequence of i.i.d. uniformly distributed random variables on $[0, 1]$, independent

of the other variables $\{\delta_j\}_{j=1}^k$ and $\{X_j, Y_j\}_{j=1}^k$. The sequence of U_j's is used to achieve randomization in the allocation. Let $\mathcal{F}_j$ be the sigma-algebra generated by $X_1, ..., X_j, Y_1, ..., Y_j, \delta_1, ..., \delta_j$, here $\mathcal{F}_0$ is the trivial sigma algebra. It is useful to consider the sigma-algebra

$$\mathcal{G}_j = \mathcal{F}_j \vee \sigma\{U_{j+1}\}. \tag{1.2}$$

where U_{j+1} is the auxiliary randomization. Hence, $\{\mathcal{G}_j, j \geq 1\}$ is an increasing sequence of sigma-algebras such that $\{X_j, Y_j\}$ is $\mathcal{G}_j$ measurable for every $j \geq 1$. Note that δ_{j+1} is $\mathcal{G}_j$ measurable and the random vector $\{X_{j+1}, Y_{j+1}\}$ is independent of $\mathcal{G}_j$. Let $N_{A,k}$ and $N_{B,k}$ denote the numbers of patients allocated to treatment A and B through stage k, then

$$N_{A,k} = \sum_{j=1}^{k} \delta_j \tag{1.3}$$

and

$$N_{B,k} = \sum_{j=1}^{k} (1 - \delta_j) = k - N_{A,k}. \tag{1.4}$$

Note that $N_{A,k}/k$, $N_{B,k}/k$ are the proportions of patients allocated to treatment A and B, respectively, by stage k. In practice, adaptive designs typically have nonzero equal initial sample sizes, so that $N_{A,k}$ and $N_{B,k}$ are not equal to zero. Also note that, due to adaptive allocation, $N_{A,k}$ and $N_{B,k}$ are random variables.

Further let $S_{A,k}$ and $S_{B,k}$ denote the numbers of successes from treatment A or B through stage k. Then

$$S_{A,k} = \sum_{j=1}^{k} \delta_j X_j \tag{1.5}$$

and

$$S_{B,k} = \sum_{j=1}^{k} (1 - \delta_j) Y_j. \tag{1.6}$$

Hence the maximum likelihood estimators of p_A and p_B are

$$\hat{p}_{A,k} = \frac{S_{A,k}}{N_{A,k}} \tag{1.7}$$

and

$$\hat{p}_{B,k} = \frac{S_{B,k}}{N_{B,k}} \tag{1.8}$$

respectively.

1.1.2 Allocation Adaptive Designs

In allocation adaptive designs, the allocation of each patient depends only on the allocations of the previous patients. These designs do not consider the response of the patients, so the individualistic issue is not addressed. A common goal of allocation adaptive designs is to achieve some degree of balance in terms of the number of patients assigned to each treatment.

Complete Randomization consists of assigning each patient to either of the two treatments with equal probability. As discussed in Efron (1971), complete randomization is used as a baseline for statistical inference while minimizing the possibility of conscious or unconscious selection bias. Sometimes, especially when the number of patients in the trial is small, complete randomization may result in some unpleasant imbalances.

The *Biased Coin Design* proposed by Efron (1971) is a modification of complete

randomization, which allocates patients to one of the two treatments according to a biased coin-tossing. Let p be a constant in $[0.5, 1)$. Let D_j denote the difference $N_{A,j}/j - N_{B,j}/j$ at stage j. Then the rule is described by:

$$P(\delta_{j+1} = 1) = \begin{cases} p & \text{if } D_j < 0 \\ 1/2 & \text{if } D_j = 0 \\ 1-p & \text{if } D_j > 0. \end{cases} \tag{1.9}$$

The allocation rule tends to balance the number of patients allocated to both treatments.

Wei (1978) noted a disadvantage of this procedure in that the allocation rule neither takes into consideration the number of patients treated thus far, nor does it discriminate between small or large absolute values of D_j. He proposed a new procedure of the biased coin type, *Adaptive Biased Coin Design*, that takes these issues into consideration. This design allocates patients according to the following rule. Let $h : [-1, 1] \to [0, 1]$ be a non-increasing function such that $h(x) = 1 - h(-x)$ for any $x \in [-1, 1]$. Then $P(\delta_{j+1} = 1) = h(D_j)$. The allocation rule will force an imbalanced experiment to be balanced in the limit.

1.1.3 Response Adaptive Designs

Recall that response adaptive designs are such that the allocation rules are based on the responses as well as the allocations of previous patients. Such designs are used when some compromise is sought between both the individualistic goal and the utilitarian goal or when some other considerations make it desirable to have unequal

numbers of patients assigned to treatments. For example, it is very natural in clinical trials to want to assign more patients to better treatment out of the two competing ones.

Zelen (1969) proposed the *Play-the-winner Rule*, where a success on one treatment results in the next patient's assignment to the same treatment, and a failure on one treatment results in the next patient's assignment to the opposite treatment.

The allocation of *Play-the-winner Rule* is deterministic, while randomness is of importance in adaptive designs. Randomization not only guards against researcher bias, but also provides probabilistic basis for an inference from the observations (See Rosenberger and Lachin, 2002). Wei and Durham (1978) incorporated randomness into the design by proposing the *Randomized Play-the-winner Rule (RPW Rule).* We consider an urn model with initial composition of μ balls of two different types, A and B. When a patient comes in, a ball is drawn and replaced. If the ball chosen is of type $i = A, B$, treatment i is assigned. The response is observed immediately, a success results in the addition of β balls of the same color and α balls of the opposite color, a failure results in the addition of β balls of the opposite color and α balls of the same color, where $\beta \geq \alpha \geq 0$. This design is denoted by $RPW(\mu, \alpha, \beta)$. Different choices of the triple (μ, α, β) give different levels of compromise between balance and allocation to the better treatment. In simulation, $RPW(1, 0, 1)$ is popular because of its simple implementation. There is one major disadvantage of $RPW(1, 0, 1)$ rule, the initial urn composition $\mu = 1$ is small. μ is an important parameter whose effect can be explored by simulation. As Rosenberger and Hu (1999) addressed, starting with just one ball of each color in the urn may lead to a higher chance that the urn

could be overwhelmed by a treatment that is very successful early on. Having a few more balls of each color to start will lead to more stable results.

So far, the designs we discussed are established with intuitive motivation, but not in terms of a target. A target is typically unknown and expressed in terms of limiting proportion, but the limit designed is motivated in different ways, e.g. precision. (See Rosenberger and Lachin, 2002). So another approach for adaptive design is based on an optimal allocation target. A large class of such rules are based on an estimate of such target by current stage. Let ν_A and $\nu_B = 1 - \nu_A$ denote the desired (limiting) allocation proportion of treatment A and B, and $\hat{\nu}_{A,j}$ and $\hat{\nu}_{B,j}$ be the estimates at stage j. Eisele (1994) and Eisele and Woodroofe (1995) introduced a *Doubly Adaptive Biased Coin Design* , where the allocation rules depend on both the current proportions on each treatment and the current estimate of desired allocation proportion. The allocation rules can be generally described by:

$$\delta_{j+1} = I\left\{U_{j+1} < \phi(\frac{N_{A,j}}{j}, \hat{\nu}_{A,j})\right\}, \tag{1.10}$$

where ϕ, the allocation function which satisfies certain regularity conditions, is a function from $[0,1]^2$ to $[0,1]$, so that the $(j+1)^{st}$ patient is allocated to A with probability $\phi(N_{A,j}/j, \hat{\nu}_{A,j})$.

Another example is the *Randomized Adaptive Design* (Melfi, Page (1998) and Melfi, Page, Geraldes (2001)), where $\phi(x,y) = y$ in this design. In the spirit of Neyman allocation, the target allocation is proposed to be $\nu_A = \sqrt{p_A q_A}/(\sqrt{p_A q_A} + \sqrt{p_B q_B})$ to minimize the variance of $\hat{p}_{A,k} - \hat{p}_{B,k}$.

Recently, Hu and Zhang (2004) modified Eisele and Woodroofe's design with

weaker and simpler conditions on the allocation function ϕ and proposed a family of allocation functions aimed at minimizing the variation of proportion $N_{A,k}/k$.

1.2 A Short Review of Resampling Methods for Adaptive Designs

This dissertation is a study of resampling methods for adaptive designs. We had a literature review of adaptive designs in the previous section. We will focus on a short review of resampling methods, which we will use for adaptive designs, in this section.

The resampling method has been applied to a variety of statistical problems and often outperformed other statistical methods, more specifically the normal approximation. Resampling methods have two major advantages:

- The resampling principle allows estimation of the sampling distribution without obtaining full knowledge of the underlying population distribution. Hence, it can be applied to any statistic, not just the sample mean. For example, if we want to estimate the variance of median, the traditional normal approximation does not work for this problem because we don't have a formula like $s/\sqrt{n}$ to provide estimated standard errors. Instead, we can calculate the variance of median from R resample observation vectors as an estimate. (See Efron and Tibshirani (1993) for more examples).

- The resampling estimate is second order correct. Singh (1981) was the first to show the second order correctness property of resampling, i.e. the rate of

resampling approximation to the sampling distribution is faster than the rate of the traditional normal approximation. In this milestone paper, he derived an almost sure Edgeworth Expansion for the distribution of the resample statistic. The work showed the resample statistic corrects the skewness of the underlying distribution and thus attains a better approximation than the normal law provides. Hence resampling estimates are more accurate compared to normal approximation.

Resampling methods are particularly useful in the context of adaptive designs. In a clinical trial, we assume patients will come in sequentially. We will allocate patients to two competing treatments A and B.

- First, note that the observation units are patients. So usually the sample size is small. For small sample size, normal approximation may not be so efficient. Resampling methods often have better results than normal approximation due to second order correctness property.

- Second, usually we are evaluating two competing treatments. Resampling methods are particularly successful when effects of treatment A and B are close. Babu (1989) illustrate this point in details by examining the second term of Edgeworth Expansions.

- Third, traditional resampling methods usually do well for i.i.d. cases, but they are not appropriate for adaptive designs. The observations of adaptive design are dependent due to adaptive allocation. We will show later in Chapter 5

that the confidence interval constructed by resampling method assuming i.i.d. responses has lower coverage probability than interval based on normal approximation. So we are looking for appropriate resampling methods to account for the dependent structure in adaptive designs.

Resampling methods for dependent data are developed as a consequence of the rapid growth of dependent data studies. Lahiri (2003) gave a review of resampling methods for dependent data. Politis, Romano and Wolf (1999) discussed subsampling, which is a special case of resampling, where the resample size is smaller than the sample size. Second order Properties were shown by Lahiri (2003) for normalized and studentized statistics under weak dependence.

In the context of adaptive design, we have some early work. Rosenberger and Hu (1999) showed for the first time some parametric resampling results in constructing confidence intervals for proportions in adaptive design. Their method does not involve resampling the original data. Instead, by computing observed estimates of the success probabilities from a clinical trial, they simulated additional trials using these estimates as the underlying probabilities. Hence, this method is called *Naive Parametric Resampling*.

In this dissertation, we will study resampling methods broadly. The major contributions of this dissertation are:

- We examine three resampling methods taking the dependence structure of adaptive design into consideration.

- Resampling consistency of resampling estimators are shown for the distribution

of sample binomial difference.

- Our simulations show that confidence intervals constructed from these resampling methods often outperform the intervals based on normal approximation.

Two basic assumptions that are in force throughout this dissertation, and will be repeated as needed, are that:

(A1) As $k \to \infty$, $N_{A,k} \to \infty, N_{B,k} \to \infty$ almost surely .

(A2) As $k \to \infty$, $N_{A,k}/k \to \nu_A$, $N_{B,k}/k \to \nu_B$ almost surely, where $\nu_A, \nu_B \in (0,1)$, and $\nu_A + \nu_B = 1$.

The rest of the paper is organized as follows. In Chapter 2, 3, and 4, we describe three proposed resampling methods for adaptive designs. They are Non-overlapping Block Bootstrap, Martingale Based Bootstrap, and Sequential Likelihood Resampling. Consistency of Non-overlapping Block Bootstrap and Martingale Based Bootstrap resampling estimators will be proved respectively. Simulation results are presented in Chapter 5. We make concluding remarks and discuss future work in Chapter 6.

Chapter 2

Non-overlapping Block Bootstrap

2.1 Introduction of Bootstrap

We will first introduce some notation for resampling methods. Let $\mathbb{Z}_k \doteq \{Z_1, ..., Z_k\}'$ be a vector of random variables with joint distribution G_k. The observed data is a realization of $\mathbb{Z}_k$. Suppose we like to estimate the population mean θ based on the observations $\mathbb{Z}_k$. Let $\hat{\theta}_k$ be the sample mean, and H_k denote the sampling distribution of the centered and scaled estimator $T_k = \sqrt{k}(\hat{\theta}_k - \theta)$. If $\{Z_1, ..., Z_k\}'$ are i.i.d. with finite mean and variance, it is well known that

$$\sqrt{k}(\hat{\theta}_k - \theta) \xrightarrow{d} N(0, \sigma^2), \tag{2.1}$$

where σ^2 is the population variance. This is so called *Normal Approximation* for sampling distribution. The statistical inference of θ, such as constructing a confidence interval for θ, is based on precise estimation of the sampling distribution H_k. Since

the joint distribution G_k is unknown, H_k remains unknown. Resampling methods typically apply to estimation for H_k.

The general procedure for resampling can be described as following:

- First, an estimator $\tilde{G}_k$ of the joint distribution G_k is constructed from the observations $\mathbb{Z}_k$.
- Second, we simulate R resample vectors, which are i.i.d. distributed as $\tilde{G}_k$. We denote the generic resample vector by $\mathbb{Z}_k^* \doteq \{Z_1^*, ..., Z_k^*\}'$, which is the sample for the resampling version of the original problem. Then we draw statistical inference for the sampling distribution H_k based on R resample vectors.

Bootstrap is a particular type of resampling method, where the resampling distribution $\tilde{G}_k$ is the product of estimators of a single marginal distribution $\tilde{F}_k$, such that

$$\tilde{G}_k = \underbrace{\tilde{F}_k \times ... \times \tilde{F}_k}_{k}. \tag{2.2}$$

Hence, the components of the resample vector $\mathbb{Z}_k^* \doteq \{Z_1^*, ..., Z_k^*\}'$ are i.i.d. $\tilde{F}_k$. A common choice of $\tilde{F}_k$ is the empirical distribution function

$$\hat{F}_k(.) \equiv k^{-1} \sum_{j=1}^{k} I(Z_j \leq .). \tag{2.3}$$

In this case, resamples are simply with replacement samples from the original observations.

The bootstrap method has been proposed in the context of dependent data. Different from i.i.d. case, the population is not characterized by the identical

marginal F only, but rather depends on the joint distribution G_k of the whole vector $\mathbb{Z}_k \doteq \{Z_1, ..., Z_k\}'$. The *Block Bootstrap* methods take care of the dependence structure by keeping the dependence within the block and taking the blocks as resampling units. For simplicity, we will focus on *Non-overlapping Block Bootstrap* and apply NBB method for estimation of p_A and p_B separately.

2.2 Weak Dependence and Stationary

Lahiri (2003) discussed in details the resampling methods for dependent data. Two basic conditions are needed for applying Non-overlapping Block Bootstrap in adaptive designs: The observation sequence is weakly dependent and strictly stationary.

Let $(X_n, n \in N)$ be a sequence of random variables. Note that these X's are general notation for introduction of definitions. *Weak Dependence* essentially says that the dependence of the process decreases as the distance m between the two segments $\{X_i : i \leq k\}$ and $\{X_i : i \geq k + m + 1\}$ increases. We first introduce the most commonly used standard measures of weak dependence: *Strong Mixing*. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let $\mathcal{A}$ and $\mathcal{B}$ be two sub σ fields of $\mathcal{F}$.

Definition 2.1: The measure for strong mixing or α mixing is given by

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|P(A \cap B) - P(A) \cdot P(B)| : A \in \mathcal{A}, \quad B \in \mathcal{B}\}. \tag{2.4}$$

Definition 2.2: Let $(X_n, n \in \mathbb{N})$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathcal{P})$. Let $\mathcal{F}_a^b = \sigma\langle\{X_i : a \leq i < b\}\rangle$, $1 \leq a \leq b \leq \infty$. The strong mixing coefficient of

$\{X_i\}_{i=1}^{\infty}$ is defined by

$$\alpha(m) = \sup\{\alpha(\mathcal{F}_1^{k+1}, \mathcal{F}_{k+m+1}^{\infty}) : k \in \mathbb{N}\}, \quad m \geq 1, \tag{2.5}$$

where $\alpha(\cdot,\cdot)$ is defined above. The process $\{X_i\}_{i\geq 1}$ is called *Strong Mixing* if $\alpha(m) \to 0$ as $m \to \infty$.

Definition 2.3: A stochastic process $\{X_t, t \in T\}$, whose index set T is linear, is said to be

(i) *strictly stationary of order k*, where k is a given positive integer, if for any k points $t_1, ..., t_k$ in T, and any h in T, the k-dimensional random vectors

$$(X(t_1), ..., X(t_k)) \quad and \quad (X(t_1 + h), ..., X(t_k + h)) \tag{2.6}$$

are identically distributed;

(ii) *strictly stationary* if for any integer k it is strictly stationary of order k.

In the context of adaptive design, since Non-overlapping Block Bootstrap will be done for $N_{A,k}$ treatment A observations and $N_{B,k}$ treatment B observations separately from the original sample $\mathbb{Z}_k$, we will check stationarity assumption for sequences of treatment A and B respectively. Let $\mathcal{F}_j$ be the sigma-algebra generated by $X_1, ..., X_j, Y_1, ..., Y_j, \delta_1, ..., \delta_j$, here $\mathcal{F}_0$ is the trivial sigma algebra. It is useful, in the proofs that follow, to consider the sigma-algebra

$$\mathcal{G}_j = \mathcal{F}_j \vee \sigma\{U_{j+1}\}, \tag{2.7}$$

where U_{j+1} is the auxiliary randomization we mentioned in Chapter 1. Hence, $\{\mathcal{G}_j, j \geq 1\}$ is an increasing sequence of sigma-algebras such that $\{X_j, Y_j\}$ is $\mathcal{G}_j$

measurable for every $j \geq 1$. Note that δ_{j+1} is $\mathcal{G}_j$ measurable and the random vector $\{X_{j+1}, Y_{j+1}\}$ is independent of $\mathcal{G}_j$. We need a theorem from Melfi and Page (2000).

Theorem 2.2.1. *Suppose that* (X_{j+1}, Y_{j+1}) *is independent of* $\mathcal{G}_j$ *for every* $j \geq 1$. *Then*

(i) $(X_1, X_2, ...)$ *are i.i.d. with common distribution* F_X*;*

(ii) $(Y_1, Y_2, ...)$ *are i.i.d. with common distribution* F_Y*;*

(iii) The above two sequences are independent of one another.

Hence, weak dependence and stationarity assumptions are both satisfied. Note that in the context of adaptive design, dependence structure is induced by adaptive allocations. Hence, $N_{A,k}$ and $N_{B,k}$ are random numbers, where $N_{A,k} + N_{B,k} = k$. Once the sequence $(X_1, X_2, ...)$ is truncated as $(X_1, X_2, ..., X_{N_{A,k}})$, the components are no longer i.i.d.. However, strong consistency and asymptotic normality for unknown parameter $\theta_X = p_A$ still hold. (See Melfi and Page (2000)). Thus, this wouldn't hurt in proving consistency of NBB estimator. We will demonstrate this in Section 2.4 in proofs of Theorem 2.4.2 and Theorem 2.4.4.

2.3 Introduction for Non-overlapping Block Bootstrap

We restrict our discussion to the case of *Non-overlapping Block Bootstrap* (NBB) method in the context of adaptive designs for estimation of p_A. Similar results will hold for estimation of p_B. Estimation of $p_A - p_B$ follows from asymptotic inde-

pendence of sample binomial difference estimator $\hat{p}_{A,k} - \hat{p}_{B,k}$. (See Melfi and Page (2000)).

First, we obtain the sample vector $\mathbb{Z}_k$, where $N_{A,k}$, $N_{B,k}$ are the numbers of observations from treatment A and B respectively. Let $\mathbb{X}_{N_{A,k}}$ denote the subgroup of $\mathbb{Z}_k$, which is composed of $N_{A,k}$ treatment A responses. Let $G_{A,k}$ denote the exact distribution of $\mathbb{X}_{N_{A,k}}$. Let $\hat{p}_{A,k} = S_{A,k}/N_{A,k}$ be an estimator of p_A based on the sample $\mathbb{X}_{N_{A,k}}$. Statistical inference of p_A is based on approximating the sampling distribution of $T_{A,N_{A,k}} = {N_{A,k}}^{1/2}(\hat{p}_{A,k} - p_A)$.

Under NBB method, the given vector of observations $\mathbb{X}_{N_{A,k}} \doteq \{X_1, ..., X_{N_{A,k}}\}'$ is partitioned into non-overlapping blocks. Let l denote the block length, b denote the total number of blocks. And suppose that l is an integer such that both l and $N_{A,k}/l$ are large, and l tends to infinity with $N_{A,k}$ but at a slower rate. For example, $l = \lfloor {N_{A,k}}^{\delta} \rfloor$ for $0 < \delta < 1$. Let $b \geq 1$ be the largest integer satisfying $lb \leq N_{A,k}$. Then, let $B_1, ..., B_b$ denote the b blocks of length l under the NBB, given by $B_1 = (X_1, ..., X_l)', ..., B_b = (X_{(b-1)l+1}, ..., X_{bl})'$. A set of b blocks are resampled with replacement from these observed blocks to generate the resample vector $\mathbb{X}_{bl}{}^* = (B_1^*, ..., B_b^*)' = \{X_1^*, ..., X_{bl}^*\}'$. Let S_{bl}^* denote the numbers of successes from resamples $\mathbb{X}_{bl^*}$ for treatment A. Let $\hat{p}_{A,k,b}$ denote the sample proportion of the first bl observations of $\mathbb{X}_{N_{A,k}}$, then

$$\hat{p}_{A,k,b} = (bl)^{-1} \sum_{j=1}^{bl} X_j, \tag{2.8}$$

which equals to $\hat{p}_{A,k}$ if $N_{A,k}$ is a multiple of l. The NBB version of $T_{A,N_{A,k}}$ is defined

as $T^*_{A,bl} \equiv \sqrt{bl}(\hat{p}^*_{A,k} - \hat{p}_{A,k,b})$, where $\hat{p}^*_{A,k} = S^*_{bl}/bl$.

The idea of Non-overlapping Block Bootstrap is: because of the strict stationarity, each block has the same l-joint distribution G_l; because of the weak dependence, these blocks are approximately independent for large values of l. Hence, we could take these blocks as approximately i.i.d. units. Let $G^b_l \equiv G_l \times ... \times G_l$, which is close to the exact joint population distribution $G_{A,k}$. By resampling from $(B_1, ..., B_b)$ randomly with replacement, the relation between $\mathbb{X}_{N_{A,k}} = \{X_1, ..., X_{N_{A,k}}\}'$ and the exact joint distribution $G_{A,k}$ can be reproduced by the relation between $\mathbb{X}^*_{bl} = (B^*_1, ..., B^*_b)' = \{X^*_1, ..., X^*_{bl}\}'$ and $\tilde{G}_l{}^b$, where $\tilde{G}^b_l$ denotes the empirical joint distribution of $(B_1, ..., B_b)'$. Let E_*, Var_* denote the expectation and variance of the resampling distribution, which are conditional on observation vector $\mathbb{X}_{N_{A,k}}$.

Let $\hat{p}_{A,k} = {N_{A,k}}^{-1} \sum_{j=1}^{N_{A,k}} X_j$ be the sample proportion. The bootstrap version is $\hat{p}^*_{A,k} = bl^{-1} \sum_{j=1}^{bl} X^*_j$.

Note that the resample blocks $\{B^*_i\}_{i=1}^b$ are i.i.d. conditional on data $\mathbb{X}_{N_{A,k}}$, with distribution

$$P_*(B^*_1 = B_i) = \frac{1}{b}, \tag{2.9}$$

for $i = 1, ..., b$.

The NBB method for estimation of p_A can be described as the following steps:

- Obtain the sample vector $\mathbb{Z}_k$, divide it into two vectors $\mathbb{X}_{N_{A,k}}$ and $\mathbb{Y}_{N_{B,k}}$, where $\mathbb{X}_{N_{A,k}}$ contains the $N_{A,k}$ treatment A responses.

- Partition $\mathbb{X}_{N_{A,k}}$ into b blocks of length l, $B_1, .., B_b$.

- Resample b blocks from $B_1, .., B_b$ with replacement.
- Calculate $\hat{p}^*_{A,k}$ from the resamples.
- Repeat steps 3 and 4 R times.
- Draw statistical inference based on R $\hat{p}^*_{A,k}$.

2.4 Theoretical Properties of NBB Estimator

2.4.1 Resampling Consistency of NBB Variance Estimator for Sample Proportions

Let $\mathbb{X}_{N_{A,k}}$ be the vector of responses from treatment A. In this section, it is shown the variance of NBB estimator $\hat{p}^*_{A,k}$ is strong consistent. Similar results extend to $\hat{p}^*_{B,k}$. Let $T_{A,N_{A,k}}$ denote the centered and scaled sample proportion, such that

$$T_{A,N_{A,k}} = \sqrt{N_{A,k}}(\hat{p}_{A,k} - p_A). \tag{2.10}$$

Suppose $b = \lfloor N_{A,k}/l \rfloor$ blocks are resampled, thus the resample size is bl. The bootstrap version of $T_{A,N_{A,k}}$ is given by

$$T^*_{A,bl} \equiv \sqrt{bl}(\hat{p}^*_{A,k} - \hat{p}_{A,k,b}). \tag{2.11}$$

The bootstrap estimator of $Var(T_{A,N_{A,k}})$ is given by $Var_*(T^*_{A,bl})$. Let $W_i = (X_{(i-1)l+1} + ... + X_{il})/l, 1 \leq i \leq b$ be the average of the i^{th} block and $W_i^* = (X^*_{(i-1)l+1} + ... + X^*_{il})/l, 1 \leq i \leq b$ be the average of the i^{th} resampled block under NBB method. Let $W_{li} = \sqrt{l}(W_i - p_A) = \sqrt{l}((X_{(i-1)l+1} + ... + X_{il})/l - p_A), 1 \leq i \leq b$ be the scaled and centered sample proportion of the i^{th} block. Note that $\hat{p}_{A,k,b} = b^{-1}\sum_{i=1}^{b} W_i$. Since $B_i^*, 1 \leq i \leq b$ are i.i.d. conditional on $\mathbb{Z}_k$ and

$$P_*(B_1^* = B_i) = \frac{1}{b}, \quad 1 \leq i \leq b. \tag{2.12}$$

We have $\hat{p}^*_{A,k} = b^{-1}\sum_{i=1}^{b} W_i^*$. Thus,

$$\begin{aligned} Var_*(T^*_{A,bl}) &= Var_*(\sqrt{bl}(\hat{p}^*_{A,k} - \hat{p}_{A,k,b})) \\ &= blVar_*((b^{-1}\textstyle\sum_{i=1}^{b}(W_i^* - p_A)) - (\hat{p}_{A,k,b} - p_A)) \\ &= l(\tfrac{1}{b}\textstyle\sum_{i=1}^{b}(W_i - p_A)^2 - (\hat{p}_{A,k,b} - p_A)^2). \end{aligned} \tag{2.13}$$

To prove consistency, we need the following lemma. Under mild moment condition, which is satisfied for Bernoulli responses, and two major assumptions we mentioned in Chapter 1, Melfi, Page and Geraldes (2001, Theorem 2) showed a central limit theorem for sample proportions, such that

Lemma 2.4.1. *As* $k \to \infty$,

$$\begin{pmatrix} \sqrt{N_{A,k}}\{\hat{p}_{A,k} - p_A\} \\ \sqrt{N_{B,k}}\{\hat{p}_{B,k} - p_B\} \end{pmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} p_A q_A & 0 \\ 0 & p_B q_B \end{bmatrix}\right). \tag{2.14}$$

Hence we have

$$\lim_{k\to\infty} Var(T_{A,N_{A,k}}) = p_A q_A. \tag{2.15}$$

To prove consistency, essentially we need to show that the bootstrap estimator $Var_*(T^*_{A,bl})$ is an estimator of the population parameter $p_A q_A$. We claim

Theorem 2.4.2. *In the context of adaptive designs, if* $l^{-1} + N_{A,k}^{-1} l = o(1)$ *as* $k \to \infty$, *then*

$$Var_*(T^*_{A,bl}) \longrightarrow p_A q_A \quad a.s., \quad as \quad k \to \infty. \tag{2.16}$$

Proof:

Recall that

$$Var_*(T^*_{A,bl}) = l(\frac{1}{b}\sum_{i=1}^{b}(W_i - p_A)^2 - (\hat{p}_{A,k,b} - p_A)^2). \tag{2.17}$$

Since $\sqrt{N_{A,k}}(\hat{p}_{A,k} - p_A) \xrightarrow{d} N(0, p_A q_A)$, in addition, $\hat{p}_{A,k,b}$ is the sample proportion of the first bl observations of $\mathbb{X}_{N_{A,k}}$, it follows that $l(\hat{p}_{A,k,b} - p_A)^2 = O(1/b) \to 0$ *a.s.*. Hence, it remains to show that

$$l\frac{1}{b}\sum_{i=1}^{b}(W_i - p_A)^2 \longrightarrow p_A q_A. \quad a.s.. \tag{2.18}$$

Note that by definition, $W_{li} = \sqrt{l}(W_i - p_A)$ is the scaled and centered sample

proportion of the i^{th} block. In addition, note that $l \to \infty$ as $k \to \infty$. Hence, for each i, $1 \leq i \leq b$,

$$W_{li} \xrightarrow{d} N(0, p_A q_A) \tag{2.19}$$

Thus,

$$E(\frac{1}{b}\sum_{i=1}^{b} W_{li}^2) \to p_A q_A, \tag{2.20}$$

Equation (2.18) holds because $(W_i - p_A)^2 \geq 0$. This finishes proof of theorem 2.4.2. □

This theorem shows the consistency of $Var_*(T^*_{A,bl})$, as long as l tends to infinity with $N_{A,k}$ but at a slower rate. There are many research work show that the optimal block length, which minimizes the asymptotic MSE of $Var_*(T^*_{A,bl})$, is of the form $l = CN_{A,k}^{1/3}(1 + o(1))$ as $N_{A,k} \to \infty$ $a.s.$, where C is a constant depends on some population parameters. For estimating the distribution function and quartiles of $T_{A,N_{A,k}}$, the optimal block length is of the form $l = CN_{A,k}^{1/4}(1 + o(1))$. (See Hall, Horowitz and Jing (1995), Lahiri (1999c, 2005)).

Our final goal is to estimate the binomial difference $p_A - p_B$. We prove the resampling consistency for $\hat{p}^*_{A,k} - \hat{p}^*_{B,k}$. Let b_A, b_B denote the total number of blocks from vectors $\mathbb{X}_{N_{A,k}}$ and $\mathbb{Y}_{N_{B,k}}$ respectively. Let $T_k^* = \sqrt{k}((\hat{p}^*_{A,k} - \hat{p}^*_{B,k}) - (\hat{p}_{A,k,b_A} - \hat{p}_{B,k,b_B}))$ be the centered and scaled resample binomial difference. We then have:

Theorem 2.4.3. *Under same conditions of Theorem 2.4.2, then*

$$Var_*(\sqrt{k}((\hat{p}^*_{A,k}-\hat{p}^*_{B,k})-(\hat{p}_{A,k,b_A}-\hat{p}_{B,k,b_B}))) \longrightarrow \frac{p_A q_A}{\nu_A}+\frac{p_B q_B}{\nu_B} \quad a.s., \quad as \quad k \to \infty. \tag{2.21}$$

Proof:

Result follows from Theorem 2.4.2, assumption 2, independent separate resampling for p_A and p_B, and Slutsky's Theorem. □

2.4.2 Resampling Consistency of NBB Estimator for the Distribution of Sample Proportions

Recall that $H_k = P(T_{A,N_{A,k}} \leq x)$ denotes the sampling distribution of $T_{A,N_{A,k}}$. Let $\hat{H}_k = P_*(T^*_{A,bl} \leq x)$ denote the distribution of $T^*_{A,bl}$. We attempt to approximate H_k by $\hat{H}_k$.

We establish the consistency of the NBB estimator for sampling distribution of $T_{A,N_{A,k}}$ by following theorem:

Theorem 2.4.4. *Under the same condition of Theorem 2.4.2,*

$$\sup_{x\in\mathbb{R}} |P_*(T^*_{A,bl} \leq x) - P(T_{A,N_{A,k}} \leq x)| \longrightarrow 0, \quad a.s. \quad as \quad k \to \infty. \tag{2.22}$$

Proof:

Since $T_{A,N_{A,k}}$ converges in distribution to $N(0, p_A q_A)$, which is continous, by

Polya's Theorem, we have

$$\sup_{x\in\mathbb{R}} |P_(T_{A,N_{A,k}} \leq x) - \Phi(x; p_A q_A)| \to 0 \quad uniformly, \quad as \quad k \to \infty. \tag{2.23}$$

Thus, it suffices to show

$$\sup_{x\in\mathbb{R}} |P_*(T^*_{A,bl} \leq x) - \Phi(x; p_A q_A)| \longrightarrow 0, \quad a.s. \quad as \quad k \to \infty. \tag{2.24}$$

Let $W_i^* = (X^*_{(l-1)i+1} + ... + X^*_{li})/l, 1 \leq i \leq b$ denote the average of the i^{th} resampled block under NBB method. Note that $W_1^*, ..., W_b^*$ are i.i.d. conditional on $\mathbb{Z}_k$, and

$$T^*_{A,bl} = \sqrt{bl}(\hat{p}^*_{A,k} - \hat{p}_{A,k,b}) = \sqrt{bl}(\frac{1}{b}\sum_{i=1}^{b} W_i^* - \hat{p}_{A,k,b}) = \sum_{i=1}^{b} \sqrt{\frac{l}{b}}(W_i^* - \hat{p}_{A,k,b}). \tag{2.25}$$

For all $\delta > 0$, let $\hat{\Delta}_{N_{A,k}} = \frac{l}{b}\sum_{i=1}^{b} E_*(W_i^* - \hat{p}_{A,k,b})^2 I(\sqrt{\frac{l}{b}}|W_i^* - \hat{p}_{A,k,b}| > \delta)$. Note that $B_i^*, 1 \leq i \leq b$ are i.i.d. conditional on $\mathbb{Z}_k$. As $b \to \infty$, by asymptotic normality of $\hat{p}_{A,k,b}$, we have

$$\begin{aligned} & \frac{l}{b}\sum_{i=1}^{b} E_*(W_i^* - \hat{p}_{A,k,b})^2 I(\sqrt{\frac{l}{b}}|W_i^* - \hat{p}_{A,k,b}| > \delta) \\ = \; & \frac{1}{b}\sum_{i=1}^{b}(W_{li} - \sqrt{l}(\hat{p}_{A,k,b} - p_A))^2 I(|W_{li} - \sqrt{l}(\hat{p}_{A,k,b} - p_A)| > b\delta) \\ = \; & (W_{l1} - \sqrt{l}(\hat{p}_{A,k,b} - p_A))^2 I(|W_{l1} - \sqrt{l}(\hat{p}_{A,k,b} - p_A)| > b\delta) \\ \to \; & 0 \quad a.s.. \end{aligned} \tag{2.26}$$

By the Central Limit Theorem for independent random variables, the distribution of $T^*_{A,bl}$ converge to $N(0, p_A q_A)$ almost surely as $k \to \infty$. This finishes the proof of Theorem 2.4.4. □

Finally, we will show the resampling consistency of NBB centered and scaled binomial difference T^*_k. Let $T_k = \sqrt{k}((\hat{p}_{A,k} - p_A) - (\hat{p}_{B,k} - p_B))$ be the centered and scaled binomial difference. We have:

Theorem 2.4.5. *Under the same condition of Theorem 2.4.2,*

$$\sup_{x \in \mathbb{R}} |P_*(T^*_k \leq x) - P(T_k \leq x)| \longrightarrow 0, \quad a.s. \quad as \quad k \to \infty. \tag{2.27}$$

Proof:

Result follows from Theorem 2.4.4, independent separate resampling for p_A and p_B, assumption 2, and Slutsky's Theorem. □

Chapter 3

Martingale Based Bootstrap

3.1 Introduction of Martingale Based Bootstrap

Martingale Based Bootstrap was introduced by Lin et al (1993) for checking the normality of Cox model. Later, Lin and Spiekerman (1996) also applied it for model checking in a parametric regression. Wang and Jing (2000) applied this method to inference for a class of functionals of survival distributions and termed it *Martingale Based Bootstrap*, abbreviated as MBB method. Recently, Wang and Wang (2001) applied it to inference for the mean difference in the two sample random censorship model.

Compared with other resamplng methods, an obvious advantage of MBB method is its simple implementation involving only resampling from a normal distribution. Suppose we want to estimate the binomial difference $p_A - p_B$. Typical resampling methods are conducted by resampling with replacement from the original sample

observations, and then calculating $\hat{p}^*_{A,k} - \hat{p}^*_{B,k}$ based on resamples as an estimate of $p_A - p_B$. Martingale Based Bootstrap follows following steps:

- First, based on the asymptotic normality of the scaled and centered sample binomial difference $\hat{p}_{A,k} - \hat{p}_{B,k}$, we construct an estimate for the asymptotic variance of $\hat{p}_{A,k} - \hat{p}_{B,k}$.
- Second, the martingale based bootstrap estimates will based on simulations from a normal distribution with mean zero and this variance estimate.

In this chapter, we will show that MBB method works well in the setting of adaptive design if we can show the martingale structure of the $\hat{p}_{A,k} - \hat{p}_{B,k}$ and prove the asymptotic normality accordingly.

3.2 Central Limit Theorem for Martingales

We first demonstrate the martingale structure of the sample binomial difference $\hat{p}_{A,k} - \hat{p}_{B,k}$. Recall from Section 1.1.1, all the moments of X_j's and Y_j's are finite because they are binary responses. We want to estimate p_A, p_B and the binomial difference $p_A - p_B$. We consider $p_A - p_B$ and note that it is easy to extend the results to p_A or p_B. In this section, it is shown that asymptotic normality of $\hat{p}_{A,k} - \hat{p}_{B,k}$ holds when the following two assumptions are satisfied.

As we mentioned in Chapter 1, two basic assumptions are in force throughout:

- Assumption 1: $N_{A,k} \to \infty, k - N_{A,k} \to \infty$ almost surely as $k \to \infty$.

- Assumption 2: $N_{A,k}/k \to \nu_A$, $N_{B,k}/k \to \nu_B$, almost surely as $k \to \infty$, where $\nu_A, \nu_B \in (0,1), \nu_A + \nu_B = 1$.

Recall that $\mathcal{F}_j$ is the sigma-algebra generated by $X_1, ..., X_j, Y_1, ..., Y_j, \delta_1, ..., \delta_j$. It is useful, in the proofs that follow, to consider the sigma-algebra

$$\mathcal{G}_j = \mathcal{F}_j \vee \sigma\{U_{j+1}\}. \tag{3.1}$$

Hence, $\{\mathcal{G}_j, j \geq 1\}$ is an increasing sequence of sigma-algebras such that (X_j, Y_j) is $\mathcal{G}_j$ measurable for every $j \geq 1$. Note that δ_{j+1} is $\mathcal{G}_j$ measurable and the random vector (X_{j+1}, Y_{j+1}) is independent of $\mathcal{G}_j$.

The sample proportions are defined by :

$$\hat{p}_{A,k} = \frac{\sum_{j=1}^{k} \delta_j X_j}{\sum_{j=1}^{k} \delta_j} \tag{3.2}$$

and

$$\hat{p}_{B,k} = \frac{\sum_{j=1}^{k} (1-\delta_j) Y_j}{\sum_{j=1}^{k} (1-\delta_j)}. \tag{3.3}$$

Melfi, Page and Geraldes (2001) proved a central limit theorem in the context of adaptive designs for general difference of sample means $\bar{X} - \bar{Y}$. Follow their spirit of proof, we have

Theorem 3.2.1. *Under assumptions 1 and 2, as $k \to \infty$,*

$$\sqrt{k}[(\hat{p}_{A,k} - \hat{p}_{B,k}) - (p_A - p_B)] \xrightarrow{d} N(0, \frac{p_A q_A}{\nu_A} + \frac{p_B q_B}{\nu_B}). \tag{3.4}$$

Proof:

Fix real constants a and b, define for each $k \geq 1$ and $j = 1, ..., k$

$$W_{kj} = (1/\sqrt{k})\{a(X_j - p_A)\delta_j + b(Y_j - p_B)(1 - \delta_j)\}. \tag{3.5}$$

Note that W_{kj} is $\mathcal{G}_j$ measurable. In addition, note that δ_j is $\mathcal{G}_{j-1}$ measurable and X_j, Y_j are independent of $\mathcal{G}_{j-1}$, then

$$\begin{aligned} E[W_{kj}|\mathcal{G}_{j-1}] &= E[(1/\sqrt{k})\{a(X_j - p_A)\delta_j + b(Y_j - p_B)(1 - \delta_j)|\mathcal{G}_{j-1}] \\ &= (1/\sqrt{k})\{a\delta_j E(X_j - p_A) + b(1 - \delta_j)E(Y_j - p_B)\} \\ &= 0. \end{aligned} \tag{3.6}$$

Hence, $\{W_{kj} : k \geq 1, 1 \leq j \leq k\}$ is a martingale difference array. Let $S_{ki} = \sum_{j=1}^{i} W_{kj}$ for $i = 1, ..., k$. Therefore, $\{S_{ki} : k \geq 1, 1 \leq i \leq k\}$ is a zero mean martingale array with differences $\{W_{kj} : k \geq 1, 1 \leq j \leq k\}$. Note that $E(S_{ki})^2 \leq a^2 + b^2 < \infty$, hence $\{S_{ki} : k \geq 1, 1 \leq i \leq k\}$ is square integrable.

By the martingale central limit theorem, see Theorem 3.2, Hall and Heyde (1980), it will follow that

$$\sum_{j=1}^{k} W_{kj} \xrightarrow{d} N(0, a^2 p_A q_A \nu_A + b^2 p_B q_B \nu_B) \tag{3.7}$$

if the following three conditions are satisfied:

$$\max_j |W_{kj}| \xrightarrow{p} 0, \tag{3.8}$$

$$\sum_j W_{kj}^2 \xrightarrow{p} a^2 p_A q_A \nu_A + b^2 p_B q_B \nu_B, \tag{3.9}$$

$$E(\max_j W_{kj}^2) \quad is \quad bounded \quad in \quad k. \tag{3.10}$$

Note that $E(S_{ki})^2 \leq a^2 + b^2$ implies condition (3.8). For condition (3.10), note that

$$\max_j |W_{kj}| \leq 1/\sqrt{k}(|a| + |b|) \xrightarrow{p} 0, \quad as \quad k \to \infty. \tag{3.11}$$

Finally, for verifying condition (3.9), note that

$$\begin{aligned} \textstyle\sum_j W_{kj}^2 - (a^2 p_A q_A \nu_A + b^2 p_B q_B \nu_B) = \; & a^2\{\tfrac{1}{k}\textstyle\sum_{j=1}^k (X_j - p_A)^2 \delta_j - p_A q_A \nu_A\} \\ & + b^2\{\tfrac{1}{k}\textstyle\sum_{j=1}^k (Y_j - p_B)^2 (1 - \delta_j) - p_B q_B \nu_B\}. \end{aligned} \tag{3.12}$$

It suffices to show that both terms on the right converge almost surely to 0.

Now, write the first term as

$$\frac{a^2}{k} \sum_{j=1}^k ((X_j - p_A)^2 - p_A q_A)\delta_j + a^2 p_A q_A (\frac{N_{A.k}}{k} - \nu_A). \tag{3.13}$$

By the assumption 2, the second term in this equation converges almost surely to 0.

For the first term, define, for each $k \geq 1$,

$$M_k = \sum_{j=1}^{k} \frac{1}{j}((X_j - p_A)^2 - p_A q_A)\delta_j. \tag{3.14}$$

Note that $\{M_k, k \geq 1\}$ is a martingale. In addition, all the moments for binary response are finite. Then we have

$$E(M_k^2) \leq E[((X_j - p_A)^2 - p_A q_A)^2] \sum_{j=1}^{k} \frac{1}{j^2} < \infty. \tag{3.15}$$

Hence, $\sup_k E(M_k^2) < \infty$. By L_2 convergence theorem, M_k converges almost surely to an almost finite limit. Kronecker's lemma implies the first term converges to 0 almost surely. The term involve b^2 in turn converges to 0 almost surely.

Let $a = 1/\nu_A$, $b = -1/\nu_B$, we have

$$W_{kj} = (1/\sqrt{k})\{((X_j - p_A)\delta_j)/\nu_A - ((Y_j - p_B)(1 - \delta_j))/\nu_B\}, \tag{3.16}$$

such that,

$$\sum_{j=1}^{k} W_{kj} \xrightarrow{d} N(0, \frac{p_A q_A}{\nu_A} + \frac{p_B q_B}{\nu_B}). \tag{3.17}$$

Note that

$$\begin{aligned} \sqrt{k}[(\hat{p}_{A,k} - \hat{p}_{B,k}) - (p_A - p_B)] \; &= \textstyle\sum_{j=1}^{k} (1/\sqrt{k})\{((X_j - p_A)\delta_j)\frac{k}{N_{A,k}} \\ &\quad -((Y_j - p_B)(1 - \delta_j))\frac{k}{N_{B,k}}\}. \end{aligned} \tag{3.18}$$

Hence, By assumption 2 and Slutsky's theorem, we finished this proof. □

This is so called normal approximation of the sampling distribution of sample binomial difference $\hat{p}_{A,k} - \hat{p}_{B,k}$. The above normal distribution can used for general statistical inference, e.g. constructing a confidence interval for $p_A - p_B$.

In Martingale Based Bootstrap method, this normal distribution can be justified as our resampling distribution. We will look at the asymptotic variance of sample binomial difference, i.e. $p_A q_A/\nu_A + p_B q_B/\nu_B$, construct an estimate of asymptotic variance, and then resample from a normal distribution with mean zero and this variance estimate. Note that, the second condition of central limit theorem for martingale, i.e. equation (3.9), provides us an natural estimator of the asymptotic variance.

By equation (3.18), replace p_A, p_B with their consistent estimators $\hat{p}_{A,k}$ and $\hat{p}_{B,k}$. Strong consistency of $\hat{p}_{A,k}$ and $\hat{p}_{B,k}$ have been shown by Melfi and Page (2000), hence the error introduced here is of $o(k^{-1/2})$, which is negligible as we calculate the variance estimator. Let

$$\hat{W}_{kj} = (1/\sqrt{k})\{((X_j - \hat{p}_{A,k})\delta_j)\frac{k}{N_{A,k}} - ((Y_j - \hat{p}_{B,k})(1-\delta_j))\frac{k}{N_{A,k}}\}. \tag{3.19}$$

Then we have,

$$\begin{aligned}\sum_{j=1}^{k} \hat{W}_{kj}^2 = & \; N_{A,k}^{-2} k \sum_{j=1}^{k} \delta_j (X_j - \hat{p}_{A,k})^2 + N_{B,k}^{-2} k \sum_{j=1}^{k} (1-\delta_j)(Y_j - \hat{p}_{B,k})^2 \\ = & \; k N_{A,k}^{-2} [(1-\hat{p}_{A,k})^2 S_{A,k} + \hat{p}_{A,k}^2 (N_{A,k} - S_{A,k})] \\ + & \; k N_{B,k}^{-2} [(1-\hat{p}_{B,k})^2 S_{B,k} + \hat{p}_{B,k}^2 (N_{B,k} - S_{B,k})].\end{aligned} \tag{3.20}$$

By Theorem 3.2.1, as $k \to \infty$, we have

$$\sum_{j=1}^{k} \hat{W}_{kj}^2 \xrightarrow{p} p_A q_A/\nu_A + p_B q_B/\nu_B. \tag{3.21}$$

From now on, we let T_k denote the centered and scaled sample binomial difference, such that

$$T_k = \sqrt{k}[(\hat{p}_{A,k} - \hat{p}_{B,k}) - (p_A - p_B)] \tag{3.22}$$

with asymptotic variance $\sigma_\infty^2 = p_A q_A/\nu_A + p_B q_B/\nu_B$.

Denote the bootstrap version of T_k by

$$T_k^* = \sqrt{k}[(\hat{p}_{A,k}^* - \hat{p}_{B,k}^*) - (\hat{p}_{A,k} - \hat{p}_{B,k})] \tag{3.23}$$

where $\hat{p}_{A,k}^* - \hat{p}_{B,k}^*$ are based on resamples from the normal distribution, i.e. $N(\hat{p}_{A,k} - \hat{p}_{B,k}, N_{A,k}^{-2}[(1-\hat{p}_{A,k})^2 S_{A,k} + \hat{p}_{A,k}^2(N_{A,k} - S_{A,k})] + N_{B,k}^{-2}[(1-\hat{p}_{B,k})^2 S_{B,k} + \hat{p}_{B,k}^2(N_{B,k} - S_{B,k})])$. In the spirit of bootstrapping, we assume the relation between sample and population can be reproduced by the relation between resample and sample. Hence, we hope that T_k^* converges to the same normal distribution as T_k does, i.e.

$$\sqrt{k}[(\hat{p}_{A,k}^* - \hat{p}_{B,k}^*) - (\hat{p}_{A,k} - \hat{p}_{B,k})] \xrightarrow{d} N(0, \frac{p_A q_A}{\nu_A} + \frac{p_B q_B}{\nu_B}). \tag{3.24}$$

We will prove this in Theorem 3.3.2.

Let ξ_k^* be a random variable from the resampling distribution which is conditional on $\mathbb{Z}_k$, such that

$$\begin{aligned} \xi_k^* \ &\sim N(0, N_{A,k}^{-2}[(1-\hat{p}_{A,k})^2 S_{A,k} + \hat{p}_{A,k}^2(N_{A,k} - S_{A,k})] \\ &+ N_{B,k}^{-2}[(1-\hat{p}_{B,k})^2 S_{B,k} + \hat{p}_{B,k}^2(N_{B,k} - S_{B,k})]). \end{aligned} \tag{3.25}$$

Clearly, $\hat{p}^*_{A,k} - \hat{p}^*_{B,k}$ can obtained as

$$\hat{p}^*_{A,k} - \hat{p}^*_{B,k} = \hat{p}_{A,k} - \hat{p}_{B,k} + \xi^*_k. \tag{3.26}$$

This will be used in the proofs in Section 3.3.

3.3 Theoretical Properties of MBB Estimator

3.3.1 Resampling Consistency of MBB Variance Estimator for the Binomial Difference

Let Var_* denote the variance of the resampling distribution, which is conditional on observation vector $\mathbb{Z}_k$. The bootstrap estimator of $Var(T_k)$ is given by $Var_*(T^*_k)$. Note that,

$$\begin{aligned} T^*_k &= \sqrt{k}[(\hat{p}^*_{A,k} - \hat{p}^*_{B,k}) - (\hat{p}_{A,k} - \hat{p}_{B,k})] \\ &= \sqrt{k}(\xi^*_k). \end{aligned} \tag{3.27}$$

Hence we show the consistency of $Var_*(T^*_k)$:

Theorem 3.3.1. *If $N_{A,k}/k \to \nu_A$ almost surely as $k \to \infty$, then*

$$Var_*(T^*_k) \xrightarrow{p} \frac{p_A q_A}{\nu_A} + \frac{p_B q_B}{\nu_B}. \tag{3.28}$$

Proof:

$$\begin{aligned} Var_*(T_k^*) &= Var_*(\sqrt{k}(\xi_k^*)) \\ &= k(Var_*(\xi_k^*)) \\ &\xrightarrow{p} \frac{p_A q_A}{\nu_A} + \frac{p_B q_B}{\nu_B}. \end{aligned} \tag{3.29}$$

□

3.3.2 Resampling Consistency of MBB Estimator for the Distribution of the Binomial Difference

Theorem 3.3.2. *Under assumptions 1 and 2,*

$$\sup_{x \in \mathbb{R}} |P_*(T_k^* \leq x) - P(T_k \leq x)| \xrightarrow{p} 0 \quad as \quad k \to \infty \tag{3.30}$$

Proof:

Since T_k converges in distribution to $N(0, \sigma_\infty^2)$, it suffices to show

$$\sup_{x \in \mathbb{R}} |P_*(T_k^* \leq x) - \Phi(x; \sigma_\infty^2)| \xrightarrow{p} 0, \quad as \quad k \to \infty. \tag{3.31}$$

Note that,

$$T_k^* = \sqrt{k}(\xi_k^*), \tag{3.32}$$

where ξ_k^* is a normal variable. It suffices to show the consistency of MBB variance

estimate for the binomial difference. Followed by Theorem 3.3.1, we have

$$Var_*(T_k^*) \xrightarrow{p} \sigma_\infty^2. \tag{3.33}$$

This finishes the proof of Theorem 3.3.2. □

Chapter 4

Sequential Likelihood Resampling

4.1 Introduction of Resampling

In this chapter we will focus on *Sequential Likelihood Resampling*, abbreviated as SLR method. Recall that, as we mentioned in Chapter 2, when the observations are dependent, the underlying joint population distribution G_k is not equal to the product of i.i.d. marginal distributions F. Let $\tilde{G}_k$ be an estimator of the joint population distribution G_k constructed from the observations $\mathbb{Z}_k \doteq \{Z_1, ..., Z_k\}'$. Resampling generalizes bootstrapping by eliminating the requirement that $\tilde{G}_k$ be the product of identical marginal estimators. It allows the estimator $\tilde{G}_k$ to be the product of a group of conditional distribution estimators based on the observations $\mathbb{Z}_k$. Resampling methods aim at capturing the underlying data generating process, which gives the dependence structure of the observations.

Recall $\mathcal{G}_j = \mathcal{F}_j \vee \sigma(U_{j+1}), 1 \leq j \leq k$ are the nested increasing sigma-algebras as we defined in Chapter 1, such that δ_{j+1} is $\mathcal{G}_j$ measurable and the random vector

$\{X_{j+1}, Y_{j+1}\}$ is independent of $\mathcal{G}_j$. Let $\mathbb{Z}_k \doteq \{Z_1, Z_2, ..., Z_k\}'$ be the vector of sample observations with joint population distribution G_k. We write the joint distribution G_k as the product of the one step ahead conditional distributions $F_j, 1 \leq j \leq k$.

$$\begin{aligned} G_k(Z_1, ..., Z_k) &= F_1(Z_1) \prod_{j=2}^{k} F_j(Z_j|\mathcal{G}_{j-1}) \\ &= \prod_{j=1}^{k} F_j. \end{aligned} \tag{4.1}$$

Let $\tilde{F}_j = \tilde{F}_j(Z_j|\mathcal{G}_{j-1}), 1 \leq j \leq k$ be an estimator of one-step ahead conditional distribution F_j. Accordingly,

$$\begin{aligned} \tilde{G}_k(Z_1, ..., Z_k) &= \tilde{F}_1(Z_1) \prod_{j=2}^{k} \tilde{F}_j(Z_j|\mathcal{G}_{j-1}) \\ &= \prod_{j=1}^{k} \tilde{F}_j. \end{aligned} \tag{4.2}$$

Let $\mathbb{Z}_k^* \doteq \{Z_1^*, ..., Z_k^*\}'$ denote a resample vector from $\{\tilde{G}_j\}_{j=1}^{k}$. Resamples are obtained by simulating from $\tilde{F}_j$ sequentially and independently, i.e.

$$Z_j^* \sim \tilde{F}_j, \qquad 1 \leq j \leq k. \tag{4.3}$$

The major idea here is that we will update the estimate of the conditional distribution F_j based on the information we obtained by stage j. The actual conditional distribution F_j is hard to estimate, we are going to construct the estimator $\tilde{F}_j$ intuitively based on the above idea. We propose to resample from a particular choice of $\tilde{F}_j$, that imposes the conditional moment restrictions.

4.2 Introduction of Sequential Likelihood Method

Sequential Likelihood Resampling method applies empirical likelihood method at each stage j, $1 \leq j \leq k$, sequentially. *Empirical Likelihood Method* was first introduced by Owen (1988). The idea of empirical likelihood is very natural and appealing. Let p_A, p_B be our parameter of interest. Let $\{Z_1, Z_2, ..., Z_k\}'$ be observations from joint distribution G_k. Let $F(Z_j)$ be the probability mass of observation Z_j. The empirical distribution function $\hat{F}_k$ assigns equal probability mass to each observation and is often considered a nonparametric maximum likelihood estimate of F_k because it maximizes the likelihood function

$$L(p_A, p_B) = \prod_{j=1}^{k} \{F(Z_j)\} \tag{4.4}$$

over all distribution functions F. Hansen in his milestone General Method of Moments (GMM) paper (1982) addressed that for small samples or dependent process, it is often advantageous to profile the maximum likelihood estimator by conditional moments which are based on current observations. The empirical likelihood method is used to numerically calculate the probability mass under linear constraints, i.e. $\sum_{j=1}^{k} \{F(Z_j)\} = 1$ and the conditional moments constraints. These are so called the *Profile Maximum Empirical Likelihood Estimators.*

With this in mind, we define the empirical likelihood function in the context of adaptive design. Note that in the context of adaptive design, we have 4 distinct outcomes:

- patients assigned to treatment A and we observe a success.

- patients assigned to treatment A and we observe a failure.
- patients assigned to treatment B and we observe a success.
- patients assigned to treatment B and we observe a failure.

At stage j, define multinomial distribution of Z_j with cell probabilities $W_j = (r_{1j}, r_{2j}, r_{3j}, r_{4j})$. Here j is fixed and r_{1j} is the probability of observing a successful treatment A, r_{2j} is the probability of observing a failed treatment A, r_{3j} is the probability of observing a successful treatment B, r_{4j} is the probability of observing a failed treatment B, such that for each $1 \leq j \leq k$, $\sum_{i=1}^{4} r_{ij} = 1$. Let $n_{ij}, i = 1, 2, 3, 4$ be the number of assignment to four cells by stage j. Note that $n_{1j} = S_{A,k}, n_{2j} = N_{A,k} - S_{A,k}, n_{3j} = S_{B,k}, n_{4j} = N_{B,k} - S_{B,k}$ in previous notation.

Our goal is to estimate p_A and p_B, which are given by

$$p_A = \frac{r_{1j}}{r_{1j} + r_{2j}}, \quad and \quad p_B = \frac{r_{3j}}{r_{3j} + r_{4j}}. \tag{4.5}$$

Note that since $\{X_j, Y_j\}$ is independent of $\mathcal{G}_{j-1}$, r_{ij}'s depend on j, but $r_{1j}/(r_{1j} + r_{2j}), r_{3j}/(r_{3j} + r_{4j})$ do not depend on j. We impose a conditional moment condition to incorporate the information from the current observations. For each $1 \leq j \leq k$, let t_j be martingale difference

$$t_j = \delta_j(X_j - p_A) + (1 - \delta_j)(Y_j - p_B). \tag{4.6}$$

We have shown in Chapter 4 that the t_j satisfy the conditional moment restriction

$$E(t_j|\mathcal{G}_{j-1}) = 0. \tag{4.7}$$

We will include this as a linear constraint as our method of profiling for finding the maximum empirical likelihood estimator (MELE) of F_j, $1 \leq j \leq k$.

Note that our observations fall into four distinct cells, accordingly let g_i denote the four possible values t_j could take, such that

$$g_1 = 1 - p_A, \quad g_2 = -p_A, \quad g_3 = 1 - p_B, \quad g_4 = -p_B. \tag{4.8}$$

Thus, for each $1 \leq j \leq k$, the profile maximum empirical likelihood estimator of $(r_{1j}, r_{2j}, r_{3j}, r_{4j})$ will maximize

$$EmpiricalLikelihood_j(p_A, p_B) = \prod_{i=1}^{4} {r_{ij}}^{n_{ij}} \tag{4.9}$$

subject to linear constraints

$$\sum_{i=1}^{4} r_{ij} = 1, \quad and \quad \sum_{i=1}^{4} g_i r_{ij} = 0. \tag{4.10}$$

Let $\tilde{F}_j = \tilde{F}_j(Z_j|\mathcal{G}_{j-1}), 1 \leq j \leq k$ be an estimator of one-step ahead conditional distribution F_j, which can be described as a multinomial distribution with cell prob-

abilities:

$$\tilde{W}_j = (\tilde{r}_{1j}, \tilde{r}_{2j}, \tilde{r}_{3j}, \tilde{r}_{4j}) = argmax_{r_{1j}, r_{2j}, r_{3j}, r_{4j}} \sum_{i=1}^{4} n_{ij} log(r_{ij}), \tag{4.11}$$

where $\sum_{i=1}^{4} r_{ij} = 1$. Numerically, the profile MELE for r_{ij}'s are given by

$$\tilde{r}_{ij} = \frac{n_{ij}}{j + \lambda g_i}, \quad 1 \leq i \leq 4. \tag{4.12}$$

where λ is the Lagrange multiplier, which solves equation

$$\sum_{i=1}^{4} \frac{g_i n_{ij}}{j + \lambda g_i} = 0. \tag{4.13}$$

Note that, $\tilde{p}_{A,j}$ and $\tilde{p}_{B,j}$ are given by

$$\tilde{p}_{A,j} = \frac{\tilde{r}_{1j}}{\tilde{r}_{1j} + \tilde{r}_{2j}}, \quad and \quad \tilde{p}_{B,j} = \frac{\tilde{r}_{3j}}{\tilde{r}_{3j} + \tilde{r}_{4j}}. \tag{4.14}$$

As a nonparametric method, the empirical likelihood method is best known for its advantage of conveniently accommodating auxiliary information containing in the adaptive allocation δ_j. See, for example, Qin and Lawless (1994, 1995), Chen and Qin (1993), and Chen and Sitter (1999).

Sequential Likelihood Resampling method can be described in some simple steps:

- Calculate $\tilde{W}_j$, $1 \leq j \leq k$, by the numerical method in equations (4.11) and (4.12) described above.

- Simulate resamples Z_j^* sequentially and independently from $\tilde{F}_j, 1 \leq j \leq k$.

- Calculate $\tilde{p}^*_{A,k}$ and $\tilde{p}^*_{B,k}$ from the resamples, where $\tilde{p}^*_{A,k} = S^*_{A,k}/N^*_{A,k}$, and $\tilde{p}^*_{B,k} = S^*_{B,k}/N^*_{B,k}$.

- Repeat steps 2 and 3 R times.

- Draw statistical inference based on R $\tilde{p}^*_{A,k}$ and $\tilde{p}^*_{B,k}$.

To prove the resampling consistency of the Sequential Likelihood Resampling estimator $\tilde{p}^*_{A,k} - \tilde{p}^*_{B,k}$, we need to solve for the Lagrange multiplier λ. The solution is complicated, see equation (4.13), so we will explore the theoretical property of SLR estimator in our future work.

Chapter 5

Simulations and Results

5.1 Constructing Confidence Intervals

Many statistical inferences can be drawn from resampling methods. In this chapter, we will focus on constructing a two-sided $100(1-\alpha)\%$ confidence interval for the binomial proportion p_A, and the binomial difference $p_A - p_B$. We want to compare the simulation results of normal approximation with each of these the resampling methods we examined in our previous chapters respectively.

There are two criteria to evaluate a confidence interval:

- *Coverage Probability:* Given a nominal confidence level, the coverage probability of the interval should be close to the nominal. Given deviance from the nominal, we prefer conservative confidence intervals rather than anti-conservative ones. Conservative means that the coverage probability of the interval is at least as large as the nominal.

- *Interval Length:* Fix the coverage probability, the shorter the interval length the better the confidence interval.

Hence, an ideal confidence interval should have at least nominal coverage probability and shorter interval length.

Many intervals based on normal approximation have been proposed. In particular, for binary responses, the *Agresti Interval* is recommended as the gold standard confidence interval for binomial proportions. See Brown, Cai and DasGupta (2001). Denote $\tilde{S}_{A,k} = S_{A,k} + 2$, $\tilde{S}_{B,k} = S_{B,k} + 2$, $\tilde{N}_{A,k} = N_{A,k} + 4$, $\tilde{N}_{B,k} = N_{B,k} + 4$. We recenter the maximum likelihood estimators $\hat{p}_{A,k}$ and $\hat{p}_{B,k}$ as:

$$\tilde{p}_{A,k} = \frac{\tilde{S}_{A,k}}{\tilde{N}_{A,k}}, \quad and \quad \tilde{p}_{B,k} = \frac{\tilde{S}_{B,k}}{\tilde{N}_{B,k}}. \tag{5.1}$$

We will compare confidence intervals constructed from our proposed three resampling methods with Agresti Interval to show the merit of resampling. In the context of adaptive designs, the procedures for constructing confidence intervals can be described as following:

Agresti Interval

- We obtain our sample vector $\mathbb{Z}_k$.

- A $100(1-\alpha)\%$ confidence interval for $p_A - p_B$ is given by

$$(\tilde{p}_{A,k} - \tilde{p}_{B,k}) \pm Z_{\alpha/2}\sqrt{(\frac{\tilde{p}_{A,k}\tilde{q}_{A,k}}{\tilde{N}_{A,k}} + \frac{\tilde{p}_{B,k}\tilde{q}_{B,k}}{\tilde{N}_{B,k}})}, \tag{5.2}$$

  where $Z_{\alpha/2}$ is the $\alpha/2$ cutoff point of a standard normal distribution.

Confidence Interval Using Resampling Methods

- Based on $\mathbb{Z}_k$, simulate from the resampling distribution R times, obtaining R sequences of treatment responses $\mathbb{Z}_k^*$.

- Compute $\hat{p}_{A,k}^* - \hat{p}_{B,k}^*$ from the resamples, these are R resampling estimates of $p_A - p_B$.

- Order these R $\hat{p}_{A,k}^* - \hat{p}_{B,k}^*$ as $(\hat{p}_{A,k}^* - \hat{p}_{B,k}^*)^{(1)}, ..., (\hat{p}_{A,k}^* - \hat{p}_{B,k}^*)^{(R)}$.

- A $100(1-\alpha)\%$ percentile confidence interval is given by

$$((\hat{p}_{A,k}^* - \hat{p}_{B,k}^*)^{(R\alpha/2)}, (\hat{p}_{A,k}^* - \hat{p}_{B,k}^*)^{(R(1-\alpha/2))}). \tag{5.3}$$

5.2 Resampling Procedure and Simulation Results

Simulations are done for various adaptive designs, such as *Randomized Play-the-winner Rule*, *Adaptive Randomized Design*, *Doubly Adaptive Weighted Design*, etc. Since the simulation results are similar, we only include here RPW(1,0,1) rule for its simple implementation. Recall that, RPW(1,0,1) rule can be described by an urn model, such that

- We have initial composition of 1 ball of each of two different colors.

- When a patient comes in, a ball is drawn and replaced.

- If the ball chosen is of color A, treatment A is assigned. If the ball chosen is of color B, treatment B is assigned.

- A success results in addition of 1 ball of the same color, a failure results in addition of 1 ball of the opposite color.

We keep allocating patients with this rule. We obtained our sample vector $\mathbb{Z}_k$. Then we will apply our three resampling methods based on this sample.

Simulations are run for a wide range of values of parameters p_A and p_B. Results are reported for p_A and $p_B = 0.1$ through 0.9 in increments of 0.1. The sample size $k = 100$, and for each sample we choose resample size $R = 2000$. We calculate coverage probability and interval length based on $N = 10000$ iterations.

Tables 5.1-5.12 show the results of coverage probability and average interval length of p_A and $p_A - p_B$ using NBB, MBB, and SLR methods. The numbers in the parentheses are coverage probabilities of Agresti method. For each combination of parameters, three dimensional plots for coverage probability and interval length of p_A and $p_A - p_B$ are included here (Figures 5.1-5.12). The dark surface indicates resampling method, the light surface indicates Agresti method.

In viewing the simulation results, it is useful to recall the numerator δ of the coefficient κ_3 of the second term of Edgeworth Expansion for sample proportions. See Brown, Cai and DasGupta (2002). We have

$$\delta = \frac{(q_A - p_A)p_A q_A}{\nu_A^2} - \frac{(q_B - p_B)p_B q_B}{\nu_B^2}. \tag{5.4}$$

In particular, κ_3 goes to zero when $p_A \sim p_B$ or $p_A \sim 1/2$, $p_B \sim 1/2$. The resampling methods will perform better in these cases.

The simulations support the expected large sample behavior. In terms of coverage probability, the resampling methods have higher coverage probability than the Agresti

Interval when p_A and p_B are close or when p_A and p_B are moderate or large. Note that this explains that resampling methods are especially useful in clinical trials, where we commonly evaluate two competing treatments, so that p_A and p_B are close.

In terms of interval length, note that NBB and SLR methods have uniformly shorter interval length than Agresti. While MBB method has similar interval length as Agresti, the difference is in the third decimal place.

It is worth noting that from the three dimensional plots, the Agresti Interval is very stable for all combinations of p_A and p_B. MBB behaves similarly to Agresti because MBB is a resampling method based on normal approximation. NBB and SLR do not behave well on the boundaries, i.e. when p_A or p_B are extremely small or extremely large. In addition, boundary behavior is different for NBB and SLR. For instance, NBB method does not do well when the difference $p_A - p_B$ is large. SLR method is bad when $p_A \sim 0$ and $p_B \sim 0$. This can be explained in terms of two factors: the nature of the design and the nature of the resampling methods.

- The nature of RPW rule. The Randomize Play-the-winner Rule, $RPW(\mu, \alpha, \beta)$ is designed to account for the ethical issue. The allocation rule will assign more patients to better treatment. If p_A or p_B are extremely small or extremely large, especially when the difference $p_A - p_B$ is large, observations from the better treatment will dominate the inferior treatment, which will cause bad inference for the inferior treatment. For example, in Table 5.1, when $p_A = 0.1$, $p_B = 0.9$, the coverage probability of p_A is low. We propose two possible solutions: (i) increase the initial urn composition. In general, starting with few more balls of

each color will lead to more stable results; (ii) increase the sample size.

- The nature of Sequential Likelihood Resampling method. SLR method is based on estimation of sequential multinomial probabilities. The Profile Maximum Empirical Likelihood Estimator is based on Maximum Likelihood Estimator with conditional moment corrections. The case when p_A and p_B are both small will lead to a higher chance that the conditional moment restriction may overcorrect the Maximum Likelihood Estimator.

5.3 Summary and Conclusion

In summary, in the context of adaptive design, under two major assumptions, although the components of observation vector $\mathbb{Z}_k$ are not i.i.d., the sample binomial difference converges to a normal distribution

$$\sqrt{k}[(\hat{p}_{A,k} - \hat{p}_{B,k}) - (p_A - p_B)] \xrightarrow{d} N(0, \frac{p_A q_A}{\nu_A} + \frac{p_B q_B}{\nu_B}). \tag{5.5}$$

Let $\hat{p}^*_{A,k}$ and $\hat{p}^*_{B,k}$ be the resampling estimators, if

$$\sqrt{k}[(\hat{p}^*_{A,k} - \hat{p}^*_{B,k}) - (\hat{p}_{A,k} - \hat{p}_{B,k})] \xrightarrow{d} N(0, \frac{p_A q_A}{\nu_A} + \frac{p_B q_B}{\nu_B}), \tag{5.6}$$

the resampling estimator is resampling consistent in distribution. Hence, the corresponding resampling method is theoretically applicable.

With this in mind, we can give a list of possible resampling methods. Note that,

the first three methods are not discussed in this dissertation:

- *I.I.D. Bootstrap.* This traditional resampling scheme is not appropriate for adaptive designs, because the treatment assignments and response are not exchangeable. The dependence structure is not accounted for. Especially, when sample size is small, the rate of convergence is too slow to lead to reasonable results. (See Tables 5.13-5.16, Figures 5.13-5.16.)

- *I.I.D. Bootstrap for p_A and p_B separately.* Dependent structure is accounted for by bootstrapping vectors of responses from treatment A and B separately.

- *Naive Parametric Resampling.* Proposed by Rosenberger and Hu (1999). the dependence structure is accounted for by simulating the adaptive rule R times using $\hat{p}_{A,k}$ and $\hat{p}_{B,k}$ as the underlying success rates. They demonstrated by simulation that this resampling method works well for adaptive designs.

- *Non-overlapping Block Bootstrap.* Blocking technique is applied. Dependency is kept within the blocks. (See Tables 5.1-5.4, Figures 5.1-5.4.)

- *Martingale Based Bootstrap.* Martingale technique is used to estimate the variance in the limit. (See Tables 5.5-5.8, Figures 5.5-5.8.)

- *Sequential Likelihood Resampling.* Dependency information is captured by resampling sequentially from a group of conditional empirical likelihood. (See Tables 5.9-5.12, Figures 5.9-5.12.)

To compare the performance of these resampling methods, we also include the

simulation results of I.I.D. Bootstrap, the only resampling method in the above list ignores the dependence structure of adaptive design.

Keep the setting of simulation same, Tables 5.13-5.16 and Figures 5.13-5.16 present the simulated coverage probability and average interval length of I.I.D. Bootstrap method comparing with Agresti method. In most cases, the coverage probability is lower than Agresti Interval. So the estimate of I.I.D. Bootstrap is not reasonable for adaptive designs.

In conclusion, there are many resampling methods that are theoretically applicable in the context of adaptive designs. Resampling methods that appropriately account for dependence structure usually will outperform the others.

Tables

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.804	0.929	0.949	0.962	0.963	0.955	0.944	0.928	0.856
	(0.969)	(0.957)	(0.953)	(0.953)	(0.947)	(0.952)	(0.951)	(0.952)	(0.955)
0.2	0.803	0.926	0.952	0.962	0.962	0.956	0.943	0.927	0.854
	(0.966)	(0.957)	(0.953)	(0.952)	(0.948)	(0.952)	(0.952)	(0.955)	(0.955)
0.3	0.795	0.925	0.951	0.963	0.963	0.956	0.944	0.924	0.848
	(0.968)	(0.959)	(0.954)	(0.954)	(0.948)	(0.952)	(0.952)	(0.953)	(0.956)
0.4	0.773	0.923	0.949	0.963	0.964	0.958	0.944	0.925	0.839
	(0.970)	(0.959)	(0.954)	(0.955)	(0.945)	(0.951)	(0.951)	(0.953)	(0.956)
0.5	0.739	0.920	0.951	0.965	0.966	0.958	0.943	0.923	0.833
	(0.972)	(0.959)	(0.956)	(0.954)	(0.946)	(0.950)	(0.949)	(0.953)	(0.956)
0.6	0.686	0.918	0.953	0.966	0.968	0.959	0.943	0.920	0.824
	(0.972)	(0.962)	(0.955)	(0.955)	(0.948)	(0.951)	(0.949)	(0.955)	(0.957)
0.7	0.610	0.911	0.955	0.967	0.968	0.958	0.942	0.916	0.807
	(0.975)	(0.964)	(0.960)	(0.955)	(0.949)	(0.950)	(0.950)	(0.952)	(0.958)
0.8	0.474	0.899	0.955	0.971	0.975	0.959	0.940	0.909	0.774
	(0.975)	(0.970)	(0.963)	(0.958)	(0.954)	(0.952)	(0.950)	(0.953)	(0.955)
0.9	0.166	0.851	0.951	0.976	0.982	0.962	0.934	0.880	0.702
	(0.972)	(0.972)	(0.968)	(0.963)	(0.957)	(0.953)	(0.948)	(0.951)	(0.957)

Table 5.1. NBB method, coverage probability of p_A

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.166	0.203	0.223	0.231	0.229	0.218	0.199	0.170	0.128
	(0.176)	(0.213)	(0.272)	(0.240)	(0.237)	(0.225)	(0.205)	(0.175)	(0.232)
0.2	0.171	0.208	0.228	0.235	0.233	0.221	0.202	0.172	0.129
	(0.182)	(0.239)	(0.238)	(0.245)	(0.242)	(0.230)	(0.209)	(0.178)	(0.134)
0.3	0.177	0.215	0.234	0.241	0.238	0.226	0.206	0.175	0.131
	(0.189)	(0.226)	(0.245)	(0.252)	(0.248)	(0.235)	(0.213)	(0.181)	(0.136)
0.4	0.185	0.222	0.242	0.249	0.245	0.232	0.211	0.179	0.133
	(0.199)	(0.236)	(0.254)	(0.261)	(0.256)	(0.242)	(0.219)	(0.178)	(0.138)
0.5	0.194	0.232	0.252	0.258	0.254	0.240	0 217	0.184	0.137
	(0.211)	(0.248)	(0.266)	(0.272)	(0.266)	(0.251)	(0.226)	(0.191)	(0.142)
0.6	0.208	0.245	0.265	0.271	0.266	0.251	0.226	0.191	0.142
	(0.228)	(0.264)	(0.282)	(0.287)	(0.280)	(0.263)	(0.236)	(0.199)	(0.148)
0.7	0.228	0.264	0.283	0.288	0.283	0.266	0.239	0.201	0.149
	(0.254)	(0.288)	(0.306)	(0.309)	(0.300)	(0.281)	(0.252)	(0.211)	(0.156)
0.8	0.259	0.292	0.310	0.314	0.307	0.288	0.259	0.217	0.161
	(0.296)	(0.327)	(0.342)	(0.343)	(0.332)	(0.310)	(0.276)	(0.231)	(0.171)
0.9	0.307	0.336	0.350	0.353	0.344	0.323	0.291	0.245	0.183
	(0.380)	(0.400)	(0.408)	(0.404)	(0.388)	(0.360)	(0.321)	(0.268)	(0.200)

Table 5.2. NBB method, interval length of p_A

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.987	0.975	0.953	0.920	0.883	0.819	0.703	0.484	0.053
	(0.982)	(0.971)	(0.961)	(0.955)	(0.954)	(0.948)	(0.941)	(0.937)	(0.931)
0.2	0.971	0.977	0.971	0.956	0.941	0.914	0.877	0.815	0.687
	(0.970)	(0.966)	(0.964)	(0.963)	(0.961)	(0.956)	(0.956)	(0.948)	(0.952)
0.3	0.952	0.972	0.972	0.965	0.954	0.945	0.927	0.907	0.886
	(0.964)	(0.963)	(0.962)	(0.958)	(0.953)	(0.956)	(0.956)	(0.953)	(0.958)
0.4	0.924	0.955	0.967	0.967	0.961	0.955	0.946	0.942	0.935
	(0.958)	(0.958)	(0.958)	(0.956)	(0.953)	(0.951)	(0.952)	(0.951)	(0.957)
0.5	0.885	0.936	0.958	0.963	0.964	0.959	0.957	0.950	0.954
	(0.953)	(0.955)	(0.957)	(0.955)	(0.954)	(0.953)	(0.953)	(0.955)	(0.956)
0.6	0.821	0.915	0.945	0.957	0.960	0.959	0.955	0.953	0.956
	(0.952)	(0.955)	(0.957)	(0.954)	(0.954)	(0.951)	(0.953)	(0.952)	(0.954)
0.7	0.703	0.876	0.928	0.947	0.956	0.956	0.952	0.948	0.945
	(0.945)	(0.954)	(0.955)	(0.951)	(0.953)	(0.951)	(0.953)	(0.953)	(0.953)
0.8	0.491	0.815	0.906	0.937	0.949	0.951	0.945	0.940	0.919
	(0.941)	(0.952)	(0.953)	(0.953)	(0.953)	(0.953)	(0.952)	(0.957)	(0.955)
0.9	0.056	0.692	0.881	0.931	0.951	0.952	0.941	0.920	0.869
	(0.938)	(0.954)	(0.957)	(0.953)	(0.951)	(0.952)	(0.951)	(0.959)	(0.966)

Table 5.3. NBB method, coverage probability of $p_A - p_B$

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.240	0.270	0.288	0.299	0.305	0.307	0.309	0.317	0.344
	(0.250)	(0.281)	(0.300)	(0.312)	(0.318)	(0.322)	(0.328)	(0.347)	(0.404)
0.2	0.269	0.298	0.316	0.326	0.332	0.334	0.337	0.345	0.367
	(0.281)	(0.310)	(0.329)	(0.341)	(0.347)	(0.351)	(0.358)	(0.375)	(0.424)
0.3	0.288	0.316	0.333	0.344	0.349	0.351	0.353	0.360	0.380
	(0.300)	(0.329)	(0.348)	(0.359)	(0.365)	(0.369)	(0.374)	(0.389)	(0.442)
0.4	0.299	0.326	0.344	0.354	0.358	0.360	0.360	0.366	0.383
	(0.312)	(0.341)	(0.359)	(0.369)	(0.374)	(0.377)	(0.381)	(0.329)	(0.430)
0.5	0.305	0.332	0.349	0.358	0.362	0.361	0.360	0.362	0.376
	(0.318)	(0.347)	(0.365)	(0.374)	(0.378)	(0.378)	(0.379)	(0.386)	(0.417)
0.6	0.307	0.334	0.351	0.360	0.361	0.358	0.354	0.351	0.360
	(0.322)	(0.351)	(0.369)	(0.377)	(0.378)	(0.375)	(0.371)	(0.372)	(0.394)
0.7	0.309	0.337	0.353	0.361	0.360	0.353	0.343	0.334	0.335
	(0.328)	(0.357)	(0.374)	(0.380)	(0.338)	(0.370)	(0.360)	(0.353)	(0.362)
0.8	0.317	0.344	0.360	0.366	0.362	0.351	0.334	0.316	0.304
	(0.346)	(0.374)	(0.389)	(0.392)	(0.386)	(0.372)	(0.352)	(0.333)	(0.327)
0.9	0.344	0.367	0.380	0.382	0.376	0.359	0.334	0.304	0.275
	(0.404)	(0.424)	(0.432)	(0.429)	(0.416)	(0.393)	(0.362)	(0.326)	(0.297)

Table 5.4. NBB method, interval length of $p_A - p_B$

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.944	0.951	0.954	0.959	0.955	0.956	0.952	0.945	0.931
	(0.966)	(0.958)	(0.954)	(0.955)	(0.95)	(0.951)	(0.949)	(0.945)	(0.94)
0.2	0.951	0.951	0.956	0.956	0.956	0.954	0.953	0.943	0.929
	(0.967)	(0.958)	(0.954)	(0.954)	(0.951)	(0.951)	(0.949)	(0.943)	(0.938)
0.3	0.953	0.952	0.956	0.956	0.953	0.955	0.953	0.944	0.925
	(0.969)	(0.958)	(0.954)	(0.952)	(0.948)	(0.951)	(0.95)	(0.944)	(0.936)
0.4	0.949	0.954	0.957	0.956	0.953	0.954	0.952	0.943	0.925
	(0.969)	(0.959)	(0.956)	(0.951)	(0.947)	(0.949)	(0.948)	(0.943)	(0.936)
0.5	0.944	0.952	0.957	0.957	0.953	0.955	0.953	0.943	0.926
	(0.972)	(0.959)	(0.955)	(0.952)	(0.945)	(0.947)	(0.947)	(0.942)	(0.936)
0.6	0.937	0.955	0.959	0.958	0.951	0.952	0.95	0.943	0.926
	(0.973)	(0.962)	(0.957)	(0.955)	(0.944)	(0.946)	(0.943)	(0.941)	(0.936)
0.7	0.94	0.956	0.96	0.96	0.955	0.953	0.95	0.941	0.923
	(0.977)	(0.966)	(0.961)	(0.956)	(0.948)	(0.944)	(0.943)	(0.939)	(0.932)
0.8	0.96	0.955	0.96	0.959	0.951	0.951	0.951	0.94	0.926
	(0.978)	(0.972)	(0.963)	(0.957)	(0.946)	(0.941)	(0.94)	(0.934)	(0.928)
0.9	0.982	0.952	0.96	0.956	0.952	0.952	0.946	0.941	0.922
	(0.976)	(0.976)	(0.971)	(0.957)	(0.947)	(0.936)	(0.932)	(0.929)	(0.923)

Table 5.5. MBB method, coverage probability of p_A

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.158	0.208	0.234	0.244	0.243	0.231	0.210	0.179	0.132
	(0.173)	(0.210)	(0.230)	(0.238)	(0.236)	(0.226)	(0.207)	(0.179)	(0.138)
0.2	0.162	0.214	0.239	0.250	0.248	0.236	0.214	0.182	0.134
	(0.179)	(0.216)	(0.235)	(0.243)	(0.241)	(0.230)	(0.211)	(0.182)	(0.140)
0.3	0.167	0.220	0.246	0.257	0.254	0.242	0.219	0.185	0.136
	(0.185)	(0.222)	(0.242)	(0.249)	(0.247)	(0.235)	(0.215)	(0.186)	(0.142)
0.4	0.173	0.228	0.255	0.265	0.263	0.249	0.225	0.190	0.139
	(0.194)	(0.231)	(0.250)	(0.257)	(0.254)	(0.242)	(0.221)	(0.190)	(0.146)
0.5	0.181	0.239	0.267	0.277	0.273	0.259	0.233	0.196	0.143
	(0.205)	(0.242)	(0.261)	(0.268)	(0.264)	(0.251)	(0.228)	(0.196)	(0.150)
0.6	0.192	0.253	0.282	0.292	0.288	0.271	0.243	0.204	0.149
	(0.220)	(0.256)	(0.275)	(0.281)	(0.277)	(0.262)	(0.238)	(0.204)	(0.156)
0.7	0.206	0.272	0.303	0.313	0.307	0.288	0.258	0.216	0.156
	(0.241)	(0.276)	(0.294)	(0.300)	(0.294)	(0.277)	(0.251)	(0.215)	(0.164)
0.8	0.227	0.298	0.333	0.343	0.336	0.315	0.280	0.233	0.169
	(0.273)	(0.306)	(0.323)	(0.326)	(0.318)	(0.300)	(0.271)	(0.232)	(0.178)
0.9	0.266	0.343	0.382	0.394	0.385	0.359	0.318	0.263	0.190
	(0.331)	(0.356)	(0.369)	(0.370)	(0.359)	(0.337)	(0.303)	(0.259)	(0.199)

Table 5.6. MBB method, interval length of p_A

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.965	0.956	0.953	0.949	0.94	0.933	0.925	0.912	0.914
	(0.981)	(0.967)	(0.96)	(0.953)	(0.942)	(0.938)	(0.93)	(0.919)	(0.924)
0.2	0.958	0.958	0.959	0.956	0.955	0.953	0.949	0.941	0.946
	(0.968)	(0.961)	(0.958)	(0.955)	(0.953)	(0.951)	(0.949)	(0.943)	(0.947)
0.3	0.951	0.957	0.96	0.956	0.957	0.958	0.956	0.951	0.948
	(0.958)	(0.957)	(0.957)	(0.953)	(0.952)	(0.952)	(0.955)	(0.95)	(0.954)
0.4	0.947	0.957	0.96	0.958	0.957	0.959	0.959	0.956	0.954
	(0.952)	(0.955)	(0.957)	(0.952)	(0.951)	(0.954)	(0.952)	(0.953)	(0.954)
0.5	0.94	0.955	0.958	0.958	0.961	0.956	0.96	0.958	0.956
	(0.944)	(0.953)	(0.953)	(0.952)	(0.953)	(0.949)	(0.953)	(0.951)	(0.951)
0.6	0.93	0.948	0.957	0.957	0.959	0.956	0.958	0.96	0.953
	(0.936)	(0.947)	(0.952)	(0.952)	(0.951)	(0.95)	(0.951)	(0.952)	(0.945)
0.7	0.927	0.948	0.955	0.957	0.957	0.955	0.954	0.959	0.954
	(0.933)	(0.947)	(0.951)	(0.951)	(0.95)	(0.948)	(0.945)	(0.953)	(0.95)
0.8	0.917	0.944	0.955	0.953	0.953	0.953	0.954	0.958	0.957
	(0.924)	(0.947)	(0.952)	(0.949)	(0.948)	(0.946)	(0.947)	(0.952)	(0.951)
0.9	0.911	0.946	0.951	0.952	0.951	0.951	0.953	0.955	0.961
	(0.925)	(0.946)	(0.955)	(0.95)	(0.947)	(0.944)	(0.946)	(0.953)	(0.961)

Table 5.7. MBB method, coverage probability of $p_A - p_B$

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.226	0.266	0.289	0.301	0.305	0.303	0.298	0.294	0.302
	(0.246)	(0.276)	(0.296)	(0.308)	(0.313)	(0.316)	(0.319)	(0.328)	(0.360)
0.2	0.266	0.303	0.326	0.339	0.345	0.347	0.347	0.351	0.371
	(0.276)	(0.305)	(0.324)	(0.336)	(0.342)	(0.345)	(0.348)	(0.357)	(0.384)
0.3	0.289	0.326	0.349	0.362	0.369	0.372	0.374	0.382	0.408
	(0.296)	(0.324)	(0.342)	(0.354)	(0.360)	(0.363)	(0.366)	(0.373)	(0.397)
0.4	0.301	0.339	0.362	0.376	0.382	0.385	0.386	0.394	0.419
	(0.308)	(0.336)	(0.354)	(0.365)	(0.370)	(0.372)	(0.373)	(0.379)	(0.398)
0.5	0.305	0.345	0.369	0.382	0.387	0.388	0.387	0.391	0.413
	(0.314)	(0.342)	(0.360)	(0.370)	(0.374)	(0.374)	(0.373)	(0.375)	(0.390)
0.6	0.304	0.347	0.372	0.385	0.388	0.385	0.380	0.378	0.391
	(0.316)	(0.345)	(0.363)	(0.372)	(0.374)	(0.372)	(0.367)	(0.365)	(0.373)
0.7	0.299	0.348	0.375	0.387	0.387	0.380	0.368	0.358	0.359
	(0.319)	(0.349)	(0.366)	(0.373)	(0.373)	(0.367)	(0.358)	(0.349)	(0.349)
0.8	0.295	0.352	0.383	0.394	0.391	0.378	0.358	0.335	0.319
	(0.329)	(0.357)	(0.373)	(0.379)	(0.376)	(0.365)	(0.349)	(0.332)	(0.319)
0.9	0.302	0.371	0.407	0.419	0.413	0.392	0.359	0.319	0.278
	(0.336)	(0.384)	(0.397)	(0.399)	(0.391)	(0.374)	(0.349)	(0.319)	(0.290)

Table 5.8. MBB method, interval length of $p_A - p_B$

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.051	0.416	0.826	0.976	0.999	0.999	0.999	0.998	0.992
	(0.968)	(0.957)	(0.952)	(0.952)	(0.949)	(0.952)	(0.950)	(0.955)	(0.957)
0.2	0.180	0.648	0.912	0.986	0.997	0.998	0.996	0.997	0.988
	(0.966)	(0.956)	(0.953)	(0.951)	(0.950)	(0.952)	(0.952)	(0.954)	(0.960)
0.3	0.340	0.788	0.946	0.988	0.996	0.996	0.994	0.992	0.983
	(0.966)	(0.960)	(0.952)	(0.953)	(0.948)	(0.953)	(0.952)	(0.953)	(0.956)
0.4	0.469	0.857	0.958	0.986	0.993	0.993	0.990	0.989	0.978
	(0.969)	(0.960)	(0.955)	(0.955)	(0.946)	(0.951)	(0.950)	(0.953)	(0.957)
0.5	0.583	0.895	0.961	0.982	0.988	0.989	0.987	0.985	0.971
	(0.970)	(0.958)	(0.954)	(0.955)	(0.947)	(0.950)	(0.951)	(0.953)	(0.957)
0.6	0.667	0.914	0.959	0.975	0.981	0.982	0.979	0.978	0.959
	(0.972)	(0.959)	(0.955)	(0.952)	(0.947)	(0.953)	(0.948)	(0.956)	(0.957)
0.7	0.733	0.922	0.953	0.967	0.973	0.973	0.971	0.970	0.948
	(0.973)	(0.962)	(0.958)	(0.951)	(0.948)	(0.949)	(0.952)	(0.955)	(0.959)
0.8	0.770	0.914	0.939	0.951	0.962	0.959	0.960	0.954	0.927
	(0.973)	(0.964)	(0.960)	(0.954)	(0.951)	(0.949)	(0.955)	(0.954)	(0.959)
0.9	0.792	0.899	0.917	0.926	0.941	0.939	0.940	0.940	0.904
	(0.974)	(0.969)	(0.962)	(0.958)	(0.954)	(0.949)	(0.951)	(0.954)	(0.961)

Table 5.9. SLR method, coverage probability of p_A

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.147	0.188	0.213	0.226	0.229	0.222	0.206	0.178	0.133
	(0.176)	(0.214)	(0.234)	(0.242)	(0.240)	(0.229)	(0.210)	(0.180)	(0.137)
0.2	0.149	0.194	0.220	0.233	0.235	0.227	0.210	0.181	0.135
	(0.181)	(0.219)	(0.239)	(0.247)	(0.245)	(0.233)	(0.213)	(0.183)	(0.139)
0.3	0.152	0.200	0.226	0.239	0.241	0.233	0.215	0.185	0.137
	(0.188)	(0.226)	(0.245)	(0.253)	(0.250)	(0.239)	(0.218)	(0.187)	(0.142)
0.4	0.156	0.205	0.233	0.245	0.247	0.238	0.220	0.189	0.140
	(0.196)	(0.233)	(0.253)	(0.260)	(0.257)	(0.245)	(0.223)	(0.191)	(0.145)
0.5	0.160	0.212	0.240	0.253	0.254	0.245	0.225	0.193	0.143
	(0.206)	(0.243)	(0.263)	(0.270)	(0.266)	(0.252)	(0.229)	(0.196)	(0.149)
0.6	0.164	0.218	0.248	0.261	0.262	0.253	0.232	0.199	0.146
	(0.218)	(0.256)	(0.275)	(0.281)	(0.277)	(0.262)	(0.238)	(0.203)	(0.154)
0.7	0.169	0.226	0.257	0.271	0.272	0.262	0.240	0.205	0.151
	(0.236)	(0.272)	(0.291)	(0.297)	(0.291)	(0.275)	(0.249)	(0.212)	(0.161)
0.8	0.173	0.234	0.267	0.282	0.283	0.272	0.250	0.213	0.156
	(0.260)	(0.295)	(0.313)	(0.317)	(0.310)	(0.292)	(0.264)	(0.224)	(0.170)
0.9	0.179	0.243	0.278	0.295	0.297	0.286	0.262	0.223	0.162
	(0.299)	(0.330)	(0.345)	(0.347)	(0.338)	(0.317)	(0.265)	(0.242)	(0.184)

Table 5.10. SLR method, interval length of p_A

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.487	0.815	0.964	0.990	0.983	0.971	0.947	0.899	0.782
	(0.983)	(0.971)	(0.961)	(0.955)	(0.954)	(0.950)	(0.943)	(0.939)	(0.930)
0.2	0.730	0.911	0.974	0.984	0.980	0.973	0.959	0.935	0.891
	(0.969)	(0.967)	(0.963)	(0.962)	(0.961)	(0.954)	(0.955)	(0.947)	(0.949)
0.3	0.826	0.937	0.974	0.980	0.977	0.972	0.964	0.948	0.925
	(0.963)	(0.961)	(0.960)	(0.956)	(0.953)	(0.953)	(0.953)	(0.955)	(0.949)
0.4	0.863	0.943	0.971	0.977	0.977	0.974	0.966	0.958	0.942
	(0.955)	(0.959)	(0.959)	(0.956)	(0.955)	(0.953)	(0.949)	(0.952)	(0.950)
0.5	0.870	0.939	0.967	0.975	0.974	0.973	0.970	0.962	0.951
	(0.953)	(0.955)	(0.959)	(0.955)	(0.957)	(0.953)	(0.954)	(0.955)	(0.951)
0.6	0.870	0.937	0.961	0.970	0.974	0.974	0.970	0.967	0.955
	(0.952)	(0.954)	(0.957)	(0.954)	(0.954)	(0.956)	(0.954)	(0.953)	(0.954)
0.7	0.845	0.918	0.947	0.960	0.967	0.967	0.968	0.965	0.956
	(0.944)	(0.952)	(0.954)	(0.951)	(0.951)	(0.953)	(0.954)	(0.957)	(0.957)
0.8	0.814	0.898	0.929	0.945	0.954	0.959	0.962	0.963	0.957
	(0.940)	(0.949)	(0.950)	(0.950)	(0.952)	(0.953)	(0.954)	(0.961)	(0.960)
0.9	0.742	0.863	0.898	0.923	0.934	0.943	0.953	0.957	0.958
	(0.936)	(0.951)	(0.947)	(0.949)	(0.952)	(0.952)	(0.955)	(0.961)	(0.972)

Table 5.11. SLR method, coverage probability of $p_A - p_B$

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.223	0.250	0.269	0.281	0.284	0.281	0.272	0.257	0.235
	(0.250)	(0.281)	(0.301)	(0.312)	(0.317)	(0.318)	(0.318)	(0.319)	(0.331)
0.2	0.255	0.284	0.303	0.315	0.319	0.317	0.310	0.298	0.281
	(0.281)	(0.310)	(0.329)	(0.340)	(0.346)	(0.347)	(0.347)	(0.350)	(0.360)
0.3	0.274	0.303	0.323	0.335	0.340	0.338	0.331	0.320	0.304
	(0.301)	(0.329)	(0.348)	(0.359)	(0.364)	(0.365)	(0.365)	(0.366)	(0.375)
0.4	0.282	0.313	0.334	0.345	0.350	0.348	0.341	0.329	0.312
	(0.312)	(0.340)	(0.358)	(0.369)	(0.374)	(0.374)	(0.373)	(0.373)	(0.379)
0.5	0.283	0.316	0.338	0.349	0.354	0.351	0.342	0.328	0.309
	(0.317)	(0.346)	(0.364)	(0.373)	(0.377)	(0.376)	(0.373)	(0.370)	(0.372)
0.6	0.277	0.313	0.336	0.348	0.352	0.348	0.337	0.320	0.297
	(0.318)	(0.347)	(0.365)	(0.374)	(0.376)	(0.373)	(0.366)	(0.359)	(0.356)
0.7	0.265	0.305	0.329	0.342	0.345	0.340	0.327	0.306	0.278
	(0.317)	(0.347)	(0.365)	(0.372)	(0.372)	(0.366)	(0.355)	(0.342)	(0.331)
0.8	0.247	0.292	0.320	0.333	0.336	0.329	0.313	0.287	0.252
	(0.319)	(0.349)	(0.366)	(0.372)	(0.369)	(0.358)	(0.342)	(0.322)	(0.301)
0.9	0.222	0.275	0.307	0.322	0.325	0.316	0.296	0.264	0.218
	(0.330)	(0.359)	(0.374)	(0.378)	(0.371)	(0.355)	(0.331)	(0.300)	(0.266)

Table 5.12. SLR method, interval length of $p_A - p_B$

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.938	0.948	0.947	0.952	0.954	0.954	0.939	0.937	0.939
	(0.955)	(0.960)	(0.950)	(0.952)	(0.956)	(0.954)	(0.943)	(0.944)	(0.959)
0.2	0.932	0.960	0.941	0.952	0.955	0.946	0.947	0.939	0.932
	(0.958)	(0.963)	(0.949)	(0.952)	(0.954)	(0.948)	(0.952)	(0.946)	(0.958)
0.3	0.939	0.953	0.949	0.948	0.953	0.945	0.948	0.940	0.936
	(0.963)	(0.961)	(0.957)	(0.951)	(0.951)	(0.943)	(0.951)	(0.947)	(0.962)
0.4	0.932	0.950	0.947	0.948	0.954	0.938	0.947	0.942	0.942
	(0.968)	(0.964)	(0.948)	(0.954)	(0.955)	(0.941)	(0.951)	(0.945)	(0.957)
0.5	0.931	0.941	0.940	0.949	0.952	0.942	0.948	0.929	0.932
	(0.962)	(0.955)	(0.942)	(0.950)	(0.954)	(0.946)	(0.949)	(0.938)	(0.960)
0.6	0.934	0.943	0.938	0.948	0.958	0.954	0.950	0.928	0.934
	(0.968)	(0.954)	(0.946)	(0.951)	(0.958)	(0.953)	(0.953)	(0.937)	(0.965)
0.7	0.920	0.944	0.943	0.945	0.954	0.949	0.954	0.932	0.937
	(0.971)	(0.958)	(0.950)	(0.950)	(0.955)	(0.951)	(0.960)	(0.934)	(0.960)
0.8	0.992	0.939	0.937	0.937	0.955	0.952	0.943	0.930	0.933
	(0.967)	(0.957)	(0.946)	(0.946)	(0.954)	(0.953)	(0.955)	(0.942)	(0.961)
0.9	0.888	0.931	0.936	0.945	0.954	0.942	0.948	0.940	0.925
	(0.970)	(0.964)	(0.954)	(0.949)	(0.951)	(0.949)	(0.960)	(0.948)	(0.955)

Table 5.13. IID Bootstrap method, coverage probability of p_A

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.110	0.145	0.163	0.170	0.168	0.160	0.145	0.122	0.088
	(0.121)	(0.152)	(0.167)	(0 173)	(0.171)	(0.163)	(0.148)	(0.126)	(0.093)
0.2	0.113	0.149	0.166	0.173	0.171	0.163	0.147	0.124	0.089
	(0.125)	(0.156)	(0.171)	(0.177)	(0.175)	(0.166)	(0.151)	(0.128)	(0.094)
0.3	0.117	0.153	0.171	0.178	0.176	0.166	0.150	0.125	0.09
	(0.129)	(0.161)	(0.176)	(0.182)	(0.179)	(0.170)	(0.154)	(0.130)	(0.096)
0.4	0.121	0.159	0.177	0.183	0.181	0.171	0.154	0.129	0.092
	(0.135)	(0.167)	(0.183)	(0.188)	(0.185)	(0.175)	(0.158)	(0.133)	(0.098)
0.5	0.126	0.166	0.184	0.191	0.188	0.177	0.159	0.132	0.094
	(0.143)	(0.175)	(0.191)	(0.196)	(0.192)	(0.181)	(0.163)	(0.137)	(0.100)
0.6	0.132	0.174	0.194	0.200	0.197	0.185	0.165	0.137	0.096
	(0.152)	(0.185)	(0.201)	(0.206)	(0.201)	(0.189)	(0.170)	(0.142)	(0.103)
0.7	0.141	0.186	0.206	0.213	0.208	0.195	0.174	0.143	0.100
	(0.166)	(0.200)	(0.216)	(0.220)	(0.214)	(0.200)	(0.179)	(0.149)	(0.108)
0.8	0.154	0.202	0.224	0.230	0.224	0.210	0.186	0.153	0.106
	(0.188)	(0.221)	(0.236)	(0.239)	(0.232)	(0.216)	(0.192)	(0.160)	(0.115)
0.9	0.173	0.228	0.252	0.257	0.250	0.232	0.204	0.167	0.114
	(0.226)	(0.258)	(0.271)	(0.271)	(0.261)	(0.241)	(0.213)	(0.176)	(0.125)

Table 5.14. IID Bootstrap method, interval length of p_A

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.957	0.951	0.947	0.946	0.939	0.937	0.937	0.921	0.862
	(0.978)	(0.962)	(0.961)	(0.957)	(0.955)	(0.952)	(0.962)	(0.961)	(0.957)
0.2	0.940	0.958	0.954	0.959	0.953	0.956	0.943	0.930	0.921
	(0.949)	(0.968)	(0.964)	(0.968)	(0.961)	(0.960)	(0.956)	(0.951)	(0.949)
0.3	0.938	0.950	0.957	0.958	0.948	0.957	0.951	0.946	0.930
	(0.954)	(0.960)	(0.957)	(0.964)	(0.954)	(0.960)	(0.955)	(0.958)	(0.949)
0.4	0.949	0.950	0.951	0.953	0.952	0.952	0.941	0.939	0.937
	(0.957)	(0.954)	(0.953)	(0.954)	(0.957)	(0.958)	(0.953)	(0.945)	(0.945)
0.5	0.929	0.949	0.951	0.948	0.936	0.944	0.942	0.933	0.940
	(0.951)	(0.957)	(0.958)	(0.957)	(0.944)	(0.950)	(0.946)	(0.938)	(0.947)
0.6	0.932	0.936	0.941	0.942	0.945	0.952	0.940	0.936	0.933
	(0.958)	(0.950)	(0.952)	(0.949)	(0.953)	(0.962)	(0.947)	(0.941)	(0.941)
0.7	0.926	0.934	0.938	0.941	0.946	0.944	0.939	0.934	0.921
	(0.954)	(0.945)	(0.946)	(0.950)	(0.956)	(0.954)	(0.945)	(0.941)	(0.936)
0.8	0.896	0.928	0.930	0.932	0.940	0.936	0.941	0.926	0.943
	(0.948)	(0.951)	(0.943)	(0.941)	(0.949)	(0.946)	(0.950)	(0.945)	(0.960)
0.9	0.854	0.920	0.928	0.931	0.935	0.940	0.938	0.938	0.929
	(0.948)	(0.953)	(0.948)	(0.942)	(0.945)	(0.949)	(0.952)	(0.958)	(0.961)

Table 5.15. IID Bootstrap method, coverage probability of $p_A - p_B$

* Numbers in the parentheses are Agresti results.

	p_A								
p_B	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.157	0.186	0.201	0.209	0.211	0.209	0.203	0.198	0.197
	(0.172)	(0.197)	(0.212)	(0.220)	(0.223)	(0.223)	(0.223)	(0.227)	(0.246)
0.2	0.185	0.211	0.227	0.235	0.238	0.239	0.237	0.238	0.246
	(0.197)	(0.221)	(0.235)	(0.244)	(0.247)	(0.249)	(0.250)	(0.256)	(0.275)
0.3	0.201	0.227	0.242	0.251	0.255	0.255	0.255	0.258	0.268
	(0.212)	(0.235)	(0.250)	(0.258)	(0.262)	(0.264)	(0.265)	(0.271)	(0.288)
0.4	0.282	0.313	0.334	0.345	0.350	0.348	0.341	0.329	0.312
	(0.220)	(0.244)	(0.258)	(0.266)	(0.270)	(0.270)	(0.271)	(0.275)	(0.290)
0.5	0.211	0.239	0.255	0.263	0.266	0.264	0.262	0.262	0.269
	(0.223)	(0.248)	(0.262)	(0.270)	(0.272)	(0.271)	(0.269)	(0.270)	(0.281)
0.6	0.209	0.239	0.256	0.263	0.265	0.262	0.256	0.251	0.253
	(0.224)	(0.249)	(0.264)	(0.270)	(0.271)	(0.268)	(0.263)	(0.259)	(0.264)
0.7	0.204	0.238	0.255	0.263	0.262	0.256	0.246	0.235	0.228
	(0.223)	(0.251)	(0.265)	(0.271)	(0.269)	(0.263)	(0.254)	(0.244)	(0.240)
0.8	0.199	0.238	0.258	0.264	0.261	0.252	0.236	0.217	0.199
	(0.227)	(0.256)	(0.270)	(0.274)	(0.270)	(0.260)	(0.244)	(0.227)	(0.212)
0.9	0.197	0.246	0.268	0.274	0.268	0.252	0.229	0.200	0.165
	(0.245)	(0.275)	(0.288)	(0.289)	(0.280)	(0.264)	(0.240)	(0.212)	(0.181)

Table 5.16. IID Bootstrap method, interval length of $p_A - p_B$

* Numbers in the parentheses are Agresti results.

Figures

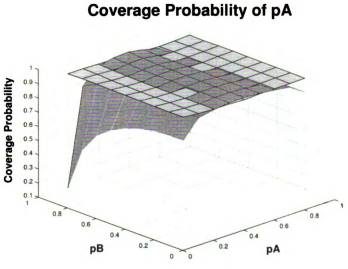

Figure 5.1. NBB method, RPW(1,0,1) rule, coverage probability of p_A.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

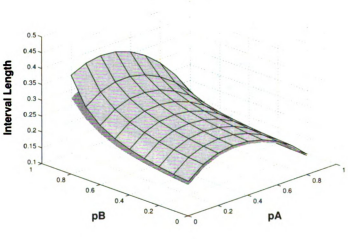

Figure 5.2. NBB method, RPW(1,0,1) rule, interval length of p_A.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

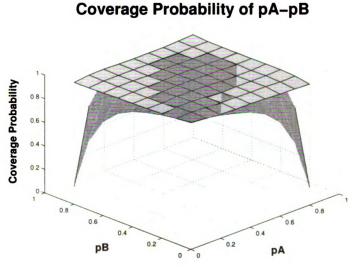

Figure 5.3. NBB method, RPW(1,0,1) rule, coverage probability of $p_A - p_B$.
$*$ The dark surface indicates resampling method,
the light surface indicates Agresti method.

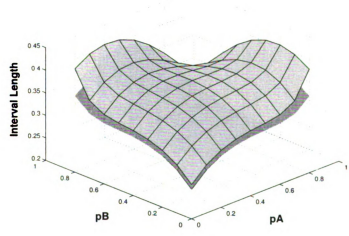

Figure 5.4. NBB method, RPW(1,0,1) rule, interval length of $p_A - p_B$.
$*$ The dark surface indicates resampling method,
the light surface indicates Agresti method.

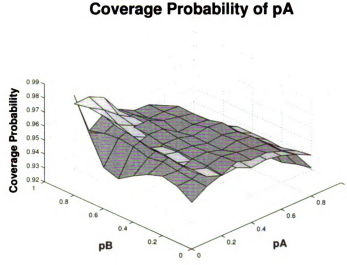

Figure 5.5. MBB method, RPW(1,0,1) rule, coverage probability of p_A.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

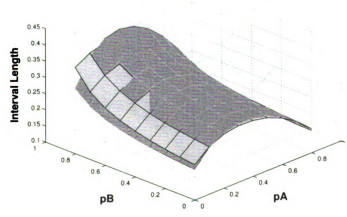

Figure 5.6. MBB method, RPW(1,0,1) rule, interval length of p_A.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

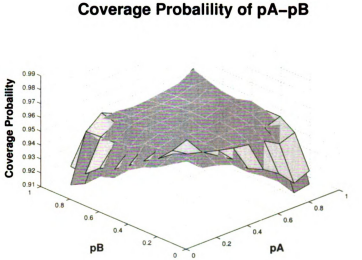

Figure 5.7. MBB method, RPW(1,0,1) rule, coverage probability of $p_A - p_B$.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

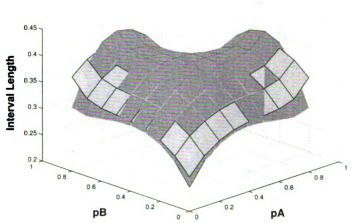

Figure 5.8. MBB method, RPW(1,0,1) rule, interval length of $p_A - p_B$.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

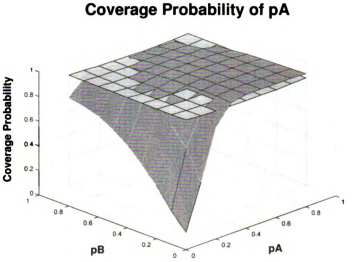

Figure 5.9. SLR method, RPW(1,0,1) rule, coverage probability of p_A.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

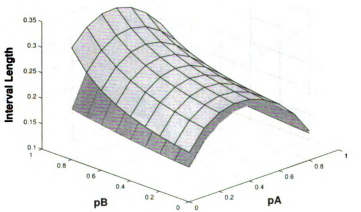

Figure 5.10. SLR method, RPW(1,0,1) rule, interval length of p_A.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

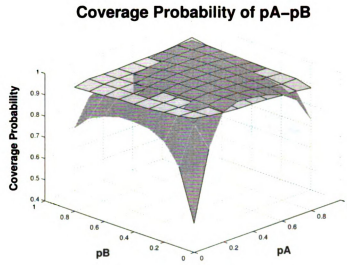

Figure 5.11. SLR method, RPW(1,0,1) rule, coverage probability of $p_A - p_B$.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

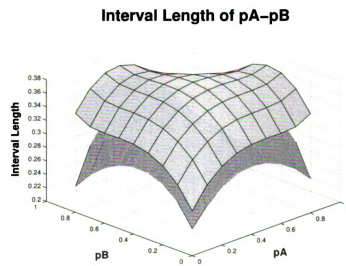

Figure 5.12. SLR method, RPW(1,0,1) rule, interval length of $p_A - p_B$.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

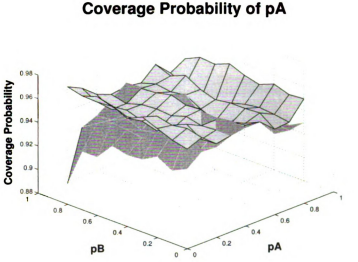

Figure 5.13. IID Bootstrap method, RPW(1,0,1) rule, coverage probability of p_A.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

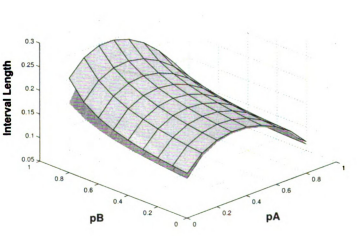

Figure 5.14. IID Bootstrap method, RPW(1,0,1) rule, interval length of p_A.
* The dark surface indicates resampling method,
the light surface indicates Agresti method.

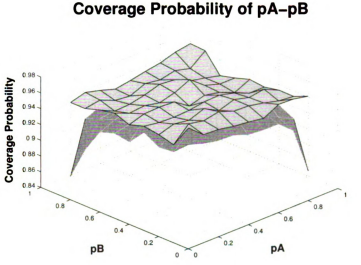

Figure 5.15. IID Bootstrap method, RPW(1,0,1) rule, coverage probability of $p_A - p_B$.
∗ The dark surface indicates resampling method,
the light surface indicates Agresti method.

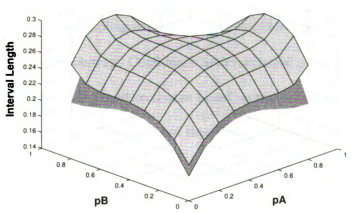

Figure 5.16. IID Bootstrap method, RPW(1,0,1) rule, interval length of $p_A - p_B$.
∗ The dark surface indicates resampling method,
the light surface indicates Agresti method.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this dissertation, we have investigated the resampling methods of adaptive design. The dependence structure is accounted for by blocking or estimating of the data generating process. The martingale structure of binomial difference was explored extensively. We proved resampling consistency for Non-Overlapping Block Bootstrap and Martingale Based Bootstrap. Confidence intervals based on resampling methods are shown to outperform the traditional ones based on normal estimation for realistic p_A and p_B in clinical trials. The results of this dissertation can be extended beyond binary response under mild assumptions of the distributions.

6.2 Future Work

Further work based on this study could take several directions. First, as we mentioned in Chapter 4, we will explore the theoretical properties of Sequential Likelihood Resampling estimator.

Second, a common tool in exploring the merit of resampling methods is the Edgeworth Expansion. The major finding is that the resampling estimator is second order correct. Edgeworth Expansion has been developed for a long time for i.i.d. observations. Basic work can be found in Hall (1988). In the context of adaptive designs, we need to go above and beyond that because the observations are dependent due to adaptive allocation. The validity of using Edgeworth Expansion needs to be checked.

Third, there are other resampling methods that could be explored in the setting of adaptive designs. The Sequential Likelihood Resampling method could be extended from simple empirical likelihood estimates to the inversion of empirical likelihood ratio test. The frame-work of empirical likelihood is natural and appealing. It is a nonparametric method but has likelihood theoretic foundations. The maximum empirical likelihood estimator is transition invariant, and a nonparametric analog of Wilks' theorem also holds: take the log-empirical likelihood ratio estimate by -2, we obtain the empirical likelihood ratio statistic (ELR) that converges to a χ^2 distribution. This is an important point, since the ELR-based test achieve asymptotic pivotalness without explicit studentization. Pivoting is theoretically important when applying bootstrap. It is often advantageous to select a pivotal statistic because the distribution of a pivotal statistic is independent of all parameters. *Implicit pivotalness*

is very useful when estimating the variance of the studentized statistic is difficult. Subsampling technique may be incorporated into the resampling. Subsampling is another branch of resampling methods, where the resample size is smaller than the original sample size. It is known that subsampling may achieve better coverage when full resampling is not.

Fourth, we use the percentile confidence interval in our simulation. There are many other non-parametric intervals can be applied in adaptive design, such as percentile-t method and BC_a method. In the BC_a method, the confidence interval incorporates the biased correction derived from Edgeworth Expansion.

Fifth, as we observed from the simulation results, the interval length is similar but not exactly the same for resampling methods vs. the Agresti Interval. We think that if we fix the interval length and then observe the coverage probability, the merit of resampling methods will be more apparent and persuasive.

Sixth, in the spirit of the empirical likelihood method, we may view some adaptive design processes as Markov Chain with four possible states. The ergodic theorem may be applied and a Monte Carlo Markov Chain (MCMC) can be used for transition probabilities. Statistical inference can be conducted based on the limiting transition probabilities.

Finally, as we mentioned in Chapter 1, in simulation, RPW(1, 0, 1) is popular because of its simple implementation. The initial urn composition is an important parameter whose effect could be explored by further simulations. We would expect more stable results by having a few more balls of each color to start with. These will be areas of further research.

Appendices

Appendices

Appendix A: Definition Index

Bibliography

Bibliography

BABU, G.J. (1989). A note on Edgeworth Expansions for the lattice case. *Journal of Multivariate Analysis* **30** 27-33.

BAI, Z.D., HU, F. and ROSENBERGER, W.F. (2002). Asymptotic properties of adaptive designs for clinical trials with delayed response. *Ann. Statist.* **30** 1-18.

BARNDORFF-NIELSEN, O.E. and COX, D.R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall.

BHATTACHARYA, R.N. and RAO, R.R. (1976). *Normal Approximation and Asymptotic Expansions.* John Wiley and Sons, Inc.

BICKEL, P.J. and DOKSUM, K.A. (2001). *Mathematical Statistics.* Prentice Hall, Inc.

BROWN, B.W. and NEWEY, W.K. (1995). *Bootstrap for GMM.* Rice University.

BROWN, B.W. and NEWEY, W.K. (1998). Efficient Semiparametric Estimation of Expectations. *Econometrica* **66** 453-464.

BROWN, B.W. and NEWEY, W.K. (2001). GMM, Efficient Bootstrapping, and Improved Inference. Manuscript.

BROWN, B.W. and NEWEY, W.K. (2002). Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference. *Journal of Business & Economic Statistics* **20, 4** 507-517.

BROWN, L.D., CAI, T.T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion. *Statistical Science* **16** 101-133.

BROWN, L.D., CAI, T.T. and DASGUPTA, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann.Statist.* **30** 160-201.

CARLSTEIN, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann.Statist.* **14** 1171-1179.

CHEN, J. and QIN, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrica* **80** 107-116.

CHEN, J. and SITTER, R.R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statist. Sinica* **9** 385-406.

CHUANG, C.S. and LAI, T.L. (1998). Resampling method for confidence intervals in group sequential trials. *Biometrica* **85** 317-332.

CHUANG, C.S. and LAI, T.L. (2000). Hybrid resampling methods for confidence intervals. *Statist. Sinica* **10** 1-50.

EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrica* **58** 404-418.

EFRON, B. and TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall.

EISELE, J. (1994). The doubly adaptive biased coin design . *Journal of Statistical Planning and Inference* **38** 249-261.

EISELE, J. and WOODROOFE, M. (1995). Central limit theorem for doubly adaptive biased coin design . *Ann.Statist.* **23** 234-254.

FLOURNOY, N. and ROSENBERGER, W.F., EDITORS (1995). *Adaptive Designs-selected proceedings of a 1992 Joint AMS-IMS-SIAM summer conference.* IMS Lecture of Notes-Monograph Series. **25**

FLOURNOY, N., ROSENBERGER, W.F., and WONG, W.K., EDITORS (1998). *New Developments and Applications in Experimental Design-selected proceedings of a 1997 Joint AMS-IMS-SIAM summer conference..* IMS Lecture of Notes-Monograph Series. **34**

HALL, P. (1985). Resampling a coverage process. *Stochastic Process Applications* **19** 259-269.

HALL, P. (1988). On the effect of random norming on the rate of convergence in the central limit theorem. *Ann. Prob.* **16** 1265-1280.

HALL, P. and HEYDE, C.C. (1980). *Martingale Limit Theory and Its Application.* Academic Press.

HALL, P., HOROWITZ, J.L. and JING, B.Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrica* **82** 561-574.

HANSEN, B.E. (1999). Non-parametric dependent data bootstrap for conditional moments models. Unpublished draft.

HANSEN, L.P. (1982). Large sample properties of generalized methods of moments estimators. *Econometrica* **50** 1029-1054.

HELMERS. R. (1982). *Edgeworth Expansions for linear combinations of order statistics.* Mathematisch Centrum, Amsterdam.

HU, F. AND ZHANG, L.X. (2004). Asymptotic Properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Ann.Statist.* **32** 268-301.

HU, F. and ZIDEK, J.V. (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika* **82** 263-275.

ISAACSON, D.L. and MADSEN, R.W. (1976). *Markov Chains.*John Wiley and Sons, Inc.

KITAMURA, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann.Statist.* **25** 2084-2102.

KITAMURA, Y., TRIPATHI, G. and AHN, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72** 1167-1714.

KUNSCH, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann.Statist.* **17** 1217-1241.

LAHIRI, S.N. (1999c). On second order properties of the stationary bootstrap method for studentized statistics, in S.Ghosh, ed., *Asymtotics, Nonparametrics, and Times Series.* Marcel Dekker, New York, pp 683-712.

LAHIRI, S.N., (2003). *Resampling Methods for Dependent Data.* Springer-Verlag, New York.

LAHIRI, S.N., (2005). Consistency of the jackknife-after-bootstrap variance estimator for the bootstrap quantiles of a Studentized statistic. *Ann.Statist.* **33** 2475-2506.

LIN, Y. and SPIEKERMAN, C.F. (1996). Model checking techniques for parametric regression with censored data. *Scand. J. Statist.* **23** 157-177.

LIN, D.Y., WEI, L.J. and YING, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrica* **80** 557-572.

MELFI, V.F. and PAGE, C. (1998). Variability in adaptive designs for estimation of success probabilities. *New developments and applications in experimental design. IMS Lecture notes-Monograph Series.* **34** 106-114.

MELFI, V.F. and PAGE, C. (2000). Estimation after adaptive allocation. *Journal of Statistical Planning and Inference* **87** 353-363.

MELFI, V.F., PAGE, C. and GERALDES, M. (2001). An adaptive randomized design with application to estimation. *Canad. J. Statist.* **29** 107-116.

NEWEY, W.K. and SMITH, R.J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72** 219-255.

OWEN, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrica* **75** 237-249.

PARZEN, E. (1962). *Stochastic Processes.* Holden-Day.

POLITIS, D.N., ROMANO, J.P. and WOLF, M. (1999). *Subsampling.* Springer-Verlag, New York.

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300-325.

QIN, J. and LAWLESS, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *Canad. J. Statist.* **23** 145-159.

ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull.Amer.Math.Soc.* **58** 527-535.

ROSENBERGER,W.F. and HU, F. (1999). Bootstrap methods for adaptive designs. *Statistics in medicine* **18** 1757-1767.

ROSENBERGER, W.F. AND LACHIN, J.M. (2002). *Randomization in clinical Trials.*John Wiley and Sons, Inc.

ROSENBERGER, W.F. and SRIRAM, T.N. (1997). Estimation for an adaptive allocation design. *Journal of Statistical Planning and Inference* **59** 309-319.

ROSENBERGER, W.F., STALLARD, N., IVANOVA, A., HARPER, C.H. and RICKS, M.L. (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57** 909-913.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187-1195.

THOMPSON, W.R. (1933). On the likelihood that one unknown probability exceeds another in the view of the evidence of the two samples. *Biometrika* **25** 275-294.

WANG, Q.H. and JING, B.Y. (2000). A martingale-based bootstrap inference with censored data. *Comm. Statist.* **29** 401-416.

WANG, Q.H. and WANG, J.L. (2001). Inference for the mean difference in the two-sample random censorship model. *Journal of Multivariate Analysis* **79** 295-315.

WEI, L.J. (1978). The Adaptive biased coin design for sequential experiments. *Ann.Statist.* **6** 92-100.

WEI, L.J. and DURHAM, S. (1978). The randomized paly-the-winner rule in medical trials. *J.Amer.Statist.Assoc.* **73** 840-843.

WEI, L.J., SMYTHE, R.T., LIN, D.Y. and PARK, T.S. (1990). Statistical inference with data-dependent treatment allocation rules. *J.Amer.Statist.Assoc.* **85** 156-162.

ZELEN, M. (1969). Play the winner rule and the controlled clinical trial. *J.Amer.Statist.Assoc.* **64** 131-146.