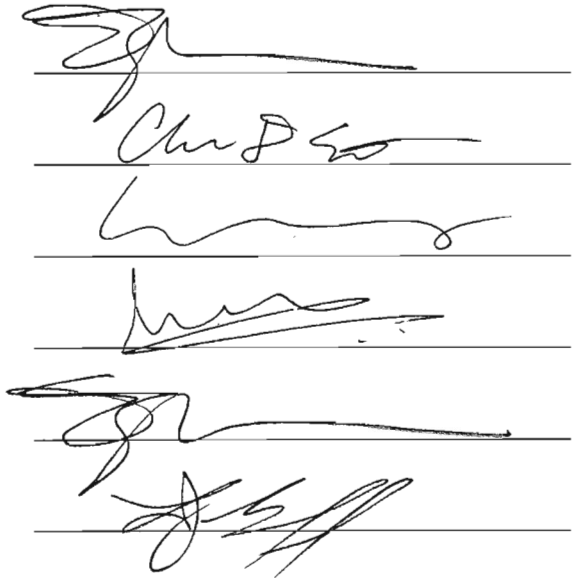
RANDOMIZATION–BASED INFERENCE FOR SEQUENTIAL CLINICAL TRIALS USING BIASED COIN RANDOMIZATION

by

Victoria Plamadeala

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Statistical Science

Committee:



Dr. William Rosenberger,
Dissertation Director

Dr. Clifton Sutton,
Committee Member

Dr. Liansheng Tang,
Committee Member

Dr. Padmanabhan Seshaiyer,
Committee Member

Dr. William Rosenberger,
Department Chair

Dr. Lloyd J. Griffiths,
Dean, The Volgenau School of
Information Technology and Engineering

Date: 11/30/2010

Fall Semester 2010
George Mason University
Fairfax, VA

Inference Using Biased Coin Randomization

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Victoria Plamadeala

Master of Science
George Mason University, 2007
Master of Arts
University of Central Florida, 2004
Licentiate
Academy of Economic Studies of Moldova, 2002

Director: Dr. William Rosenberger, Professor and Chairman
Department of Statistics

Fall Semester 2010
George Mason University
Fairfax, VA

Dedicată mamei mele, Ana Plămădeală, din tot sufletul

Acknowledgments

The completion of this dissertation would not have been possible without the help and support of the following people.

I am sincerely grateful to my supervisor, Dr. Rosenberger, for clearly defining, and allowing me to work on, the problems addressed in this thesis. I appreciate his guidance through the development of the solutions, while giving me the independence to work on my own. I thank him for the chance to give my first talk at a statistics conference and for exposing me to the research community in statistics. I thank him for trusting my ability to grow despite my many weaknesses and shortcomings as a statistics student.

I am immensely appreciative to Dr. Sutton for teaching me statistics and for having to endure many senseless answers in my homeworks and exams during my beginning years at GMU. He has influenced the way I think not only by being my professor in five courses but also by mentoring me in the Statistical Consulting Center. I also thank him for serving on the committee of this thesis.

My heartfelt gratitude goes to Liz Quigley for helping me schedule numerous meetings, reserve conference rooms, decide on the proper forms to be filled, find my lost umbrella – always with a genuine smile.

I thank Dr. Fiefang Hu of University of Virginia for helpful suggestions on this work and Drs. Tang and Seshaiyer for being committee members for this dissertation.

I thank Alastair Neil and Jonathan Goldman of the Volgenau School computing services and Drs. Gantz and Miller for accommodating the computing needs for my simulations.

I thank Dr. Junsoo Lee of University of Alabama, formerly at University of Central Florida, for inspiring me to have an interest in statistics. I still wonder if I'd ever taken this path had a different person taught research methods in economics in the fall of 2002 at UCF.

I remain indebted to my family for having valued education and self–improvement even in virtually impossible financial circumstances, and to my husband, Geoff, for his unconditional support.

Table of Contents

List of Tables

List of Figures

Abstract

INFERENCE USING BIASED COIN RANDOMIZATION
Victoria Plamadeala, PhD
George Mason University, 2010
Dissertation Director: Dr. William Rosenberger

We provide a novel approach to approximate conditional randomization tests following Efron's randomization procedure by sampling from the conditional reference set. We use combinatorial algebra to derive the conditional distribution of the number of subjects randomized to a treatment. The result is a simple and efficient Monte Carlo technique that is invariant to the total sample size, the degree of imbalance between treatments, the choice of test statistic, or the biased coin parameter. Moreover, it provides an unbiased and strongly consistent estimator for the conditional randomization test p–value. Additionally, the technique is easily extended to the approximation of conditional *stratified* randomization tests. Finally, sampling from the conditional reference set enables the approximation of conditional randomization tests when *sequential monitoring* is performed in the course of the experiment.

Chapter 1: Introduction

1.1 Background and literature review

A clinical trial is a multi–phase experiment "designed to evaluate the beneficial and adverse effects of a new medical treatment or intervention" (Rosenberger and Lachin, 2002, p. 1). In a phase III clinical trial, a new therapy is compared to a conventional therapy or placebo by randomly assigning patients to treatments (Rosenberger and Lachin, 2002, pp. 6–7). The purposes of randomization are to ensure comparability among the treatment groups with respect to both known and unknown covariates and to minimize certain biases, including selection bias (Rosenberger and Lachin, 2002, p. 7). Besides serving these purposes, randomization also provides a basis for hypothesis testing. The latter role is at the focus of this thesis.

Although randomization can be done to multiple treatment groups, this thesis addresses the case of two treatments only. Let $\boldsymbol{T} = (T_1, ..., T_n)'$ be a randomization sequence, where $T_i = 1$ if subject i is randomized to treatment 1; $T_i = 0$ if treatment 2, $i = 1, ..., n$. Let $N_1(j) = \sum_{i=1}^{j} T_i$, the number of subjects randomized to treatment 1 after j assignments, and let $D_j = 2N_1(j) - j$, the difference in the number of subjects assigned to treatments 1 and 2. A *restricted* randomization procedure is given by $\phi_{j+1} = P(T_{j+1} = 1|D_j)$, and is used to balance the number of subjects in each treatment. If $T_1, ..., T_n$ are i.i.d. Bernoulli with parameter $P(T_i = 1) = 1/2,\ i = 1, ..., n$, randomization is *complete*. The main properties of $\boldsymbol{T}$ are determined by the variance–covariance matrix of $\boldsymbol{T}$.

Efron's (1971) famous biased coin design is a restricted randomization procedure for clinical trials that has exceptional properties: it balances treatment assignments throughout the course of the trial with low variability (e.g., Baldi Antognini, 2008), and it mitigates selection and accidental biases (Rosenberger and Lachin, 2002). The biased coin design BCD(p) for a parameter $p \in [1/2, 1]$, is defined as:

$$\phi_{j+1} = \begin{cases} 1/2, & \text{when} \quad D_j = 0, \\ p, & \text{when} \quad D_j < 0, \quad j = 0, 1, 2, ... \\ 1-p, & \text{when} \quad D_j > 0. \end{cases}$$

When $p = 0.5$, randomization is complete. When $p = 1$, the randomization corresponds to a permuted block design with block size 2. When $p < 1$, the design is *fully randomized*, in that each subject will be assigned to a treatment randomly, which differs markedly from the permuted block design, where some subjects in the tail of each block are assigned with probability 1. The sequence $\{|D_n|\}_{n=1}^{\infty}$ forms a Markov chain of period 2 with states $0, 1, 2, ...$, and a reflecting barrier at the origin, and has stationary probabilities π_j given by

$$\pi_0 = \frac{p/(1-p) - 1}{2p/(1-p)}, \; \pi_j = \frac{\{p/(1-p)\}^2 - 1}{2\{p/(1-p)\}^{j+1}} \;\; (j \geq 1)$$

(Efron, 1971). Markaryan and Rosenberger (2010) derive the exact distribution of D_n as well as the exact expression for the variance–covariance matrix of $\boldsymbol{T}$ under the BCD(p).

Efron's BCD(p) is a special case of a more general framework of restricted randomization procedures called *generalized biased coin designs* (GBCD), put forward

by Wei (1978). This procedure has the form $\phi_{j+1} = \zeta\left(D_j/j\right)$, where $\zeta(\cdot)$ is a non-increasing function satisfying $\zeta(x) + \zeta(-x) = 1$. (The BCD(p) is shown to be a special case of the GBCD by letting $\zeta(x) = p$, for $-1 \leq x \leq 0$, $\zeta(0) = 1/2$, and $\zeta(x) = 1 - p$, for $0 < x \leq 1$.) The main result in Wei (1978) is showing that if $\zeta(x)$ is differentiable at $x = 0$, $n^{-1/2}D_j$ converges to a normal distribution with mean 0 and variance $\{1 - 4\zeta'(0)\}^{-1}$, where $\zeta'(0)$ is the first derivative of $\zeta(x)$ at $x = 0$. Although this result does not apply to the BCD(p), since $\zeta(x)$ in the case of the BCD(p) is not continuous, it does apply to the following class of designs proposed by Smith (1984):

$$\phi_{j+1} = \frac{\{j - N_1(j)\}^\rho}{\{N_1(j)\}^\rho + \{j - N_1(j)\}^\rho},$$

where ρ is a positive parameter. Exact results pertaining to the distribution of D_j under the design by Smith (1984) are unknown at present.

One of the significant properties of randomized designs is its natural structure induces a basis for inference. While clinical trials do not follow the usual random sampling from a population basis for statistical inference, they are usually analyzed as though there were such a basis. Such an invoked model ignores the only random component of a clinical trial design – the randomization procedure itself – which can be used to build a valid hypothesis test, named the *randomization test*. Let $\boldsymbol{z} = (z_1, ..., z_n)' = (x_1, ..., x_{n_1}; y_1, ..., y_{n-n_1})'$ be responses obtained after randomly assigning a *nonrandom* sample of n subjects to two treatments: n_1 subjects being randomly assigned to treatment 1 and the remaining $n - n_1$ to treatment 2 (Lehmann, 1986, p. 237). The only random element is the random assignment rule $\boldsymbol{T}$ (Lehmann, 1986, pp. 245–246). A valid level–α test about the hypothesis of no treatment effect can be built based only on $\boldsymbol{T}$. Under the hypothesis of no treatment effect, the z_i's

are deemed to be an arrangement of responses unaffected by treatment, that would have turned out to be the same regardless of the observed randomization sequence $\boldsymbol{t}$ and so, any apparent treatment effect is solely due to the random assignment $\boldsymbol{T}$.

To measure the treatment effect, a metric must be established. While any metric depicting the treatment effect can be used with randomization–based inference, the family of *linear rank tests* provides a large class of test statistics with which to conduct randomization tests. The form of the statistic is $S(\boldsymbol{T}) = \boldsymbol{a}_n'\boldsymbol{T}$, for a score vector $\boldsymbol{a}_n = (a_{1n} - \bar{a}_n, ..., a_{nn} - \bar{a}_n)'$, where a_{jn} is some function of the ranks of the z_i's and $\bar{a}_n = \sum_{j=1}^{n} a_{jn}/n$. For a clinical trial with binary outcomes, using binary scores where $a_{jn} = 1$ or 0, yields the randomization version of the Pearson's chi–square test. For a clinical trial with continuous outcomes, using *simple rank scores*, given by $a_{jn} = r_{jn}$, where r_{jn} are the integer ranks, leads to the randomization–based version of the Wilcoxon rank–sum test. For survival data, using *Savage scores*, given by $a_{jn} = E(X_{(j)}) - 1$, where $X_{(1)}, ..., X_{(n)}$ are order statistics from unit exponential random variables, yields the randomization-based version of the logrank test when there are no ties or censoring.

The null hypothesis in a randomization test has the form of "no group or treatment difference". The alternative hypothesis can either be general – "the groups are different", which leads to a two–sided test, or a statement about a positive or negative group difference, which leads to a one–sided test. When testing the hypothesis of no treatment differences, the randomization test p–value is obtained by adding the probabilities of sequences that yield a more extreme statistic than the observed sequence in the direction of the alternative hypothesis. Let Ω_u be the set of all possible sequences of 0's and 1's, called the *unconditional reference set* (Rosenberger and Lachin, 2002, pp. 95–96). When one generates all possible randomization sequences and computes the p–value based on the unconditional reference set, such

a test is called an *unconditional test*. More formally, the p–value of an upper tailed unconditional randomization test is obtained as

$$\sum_{\boldsymbol{t}\in\Omega_u} I(S(\boldsymbol{t}) \geq S_{obs.})P(\boldsymbol{T} = \boldsymbol{t}), \tag{1.1}$$

where $I(\cdot)$ is the indicator function and $S_{obs.}$ is the observed value of the test statistic. However, the unconditional reference set contains sequences that give little or no information about the treatment effect (e.g., $1, 1, \cdots, 1$). For that reason, one can condition $\boldsymbol{T}$ on $N_1(n)$, i.e., the observed number of subjects on treatment 1 after n subjects were randomized, and compute the upper tailed p–value of the test as

$$\sum_{\boldsymbol{t}\in\Omega_c} I(S(\boldsymbol{t}) \geq S_{obs.})P(\boldsymbol{T} = \boldsymbol{t}|N_1(n) = n_1), \tag{1.2}$$

where $\Omega_c \subset \Omega_u$ is the set of sequences that satisfy $N_1(n) = n_1$. This leads to a *conditional test*. Cox (1982) suggested that conditioning on only sequences that have $N_1(n) = n_1$ may be somewhat restrictive, and that expanding the reference set to include sequences close to n_1 but not exactly equal to n_1 might be a better approach. We refer to this approach as the "quasi–conditional" test, and examine sequences where $N_1(n) = n_1 \pm \gamma n$ for some small $\gamma > 0$ (e.g., 0.05).

For restricted randomization procedures, the conditional distribution of $S(\boldsymbol{T})$ can be asymmetric, so doubling the one–sided p–value is not appropriate when computing a two–sided p–value. Note that $\mu_{|n_1} = E(S(\boldsymbol{T})|N_1(n) = n_1)$ is not necessarily zero, even with centered scores. For exact tests, Cytel Inc. (2007, p. 509) recommends computing two-sided p–values by augmenting the one–sided p–value with the corresponding opposite tail, at an equal distance from the mean. In this case, the two–sided test p–value is given by $P\left(|S(\boldsymbol{T}) - \mu_{|n_1}| \geq |S_{obs} - \mu_{|n_1}\right\| N_1(n) = n_1\big)$,

which is the approach we take in computing the two–sided p–value of conditional randomization tests.

With a given vector of observations, test statistic and type of reference set (conditional or unconditional), the p–value of the test depends further on $\boldsymbol{T}$. Mehta, Patel and Wei (1988) noticed, by means of simulation, that with mild to severe time trends in the observations, the randomization distributions of the Wilcoxon two–sample statistic based on different randomization procedures varied nontrivially. They also conjecture that if the test p–values computed under complete randomization and under a restricted randomization procedure differed noticeably, it may be an indication that the population model assumption is invalid.

Most expositions of randomization tests in the literature also mention *permutation tests*. Some authors use these names interchangeably (e.g., Good, 2000, p. 269, Kalish and Begg, 1987), while others stress their differences (e.g., Edgington and Onghena, 2007, p. 288; Kempthorne, 1986, p. 524). Below we attempt to clarify in what sense these tests are different and similar. The following presentation of permutation tests is based on Lehmann (1986), Shao (2003) and Pesarin (2001). Let $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n-n_1}$ be random and independent samples from continuous distributions F_X and F_Y, respectively. Let $\boldsymbol{z}$ be a realization of $\boldsymbol{Z} = (Z_1, ..., Z_n)' = (X_1, ..., X_{n_1}; Y_1, ..., Y_{n-n_1})'$. Also, let $S(\boldsymbol{z})$ be the set of all points in $\Re^n$ obtained by permuting $\boldsymbol{z}$ in all $n!$ ways. A level–α test φ for $H_0 : F_X = F_Y$ vs. $H_1 : F_X \neq F_Y$ that satisfies

$$\frac{1}{n!} \sum_{\boldsymbol{z}^* \in S(\boldsymbol{z})} \varphi(\boldsymbol{z}^*) = \alpha \tag{1.3}$$

is called a *permutation test* (Lehmann, 1986, pp. 230–231). Condition (1.3) is a consequence of using the sufficiency of the set of order statistics $T(\boldsymbol{Z}) = (Z_{(1)}, ..., Z_{(n)})'$

for F when testing $H_0 : F_X = F_Y = F$. Under the null, $Z_1, ..., Z_n \overset{\text{iid}}{\sim} F$ and the set of order statistics $T(\boldsymbol{Z})$ is a sufficient statistic for F, since the sample space of $T(\boldsymbol{Z})$ is Euclidean and the conditional probability distribution of $\boldsymbol{Z}$ given $T(\boldsymbol{Z}) = \boldsymbol{z}_{(1:n)}$ (the vector of observed order statistics) does not depend on F. In fact, this conditional distribution is

$$P\{\boldsymbol{Z} = \boldsymbol{z}^* | T(\boldsymbol{Z}) = \boldsymbol{z}_{(1:n)}\} = \begin{cases} \frac{1}{n!}, & \text{when } \boldsymbol{z}^* \in S(\boldsymbol{z}), \\ 0, & \text{otherwise.} \end{cases} \tag{1.4}$$

Using (1.4),

$$E_0\{\varphi(\boldsymbol{Z}) | T(\boldsymbol{Z}) = \boldsymbol{z}_{(1:n)}\} = \frac{1}{n!} \sum_{\boldsymbol{z}^* \in S(\boldsymbol{z})} \varphi(\boldsymbol{z}^*),$$

which is the size of the test given in (1.3). Furthermore, the distribution of the test statistic in a permutation test follows from (1.4). This distribution is called the permutation distribution.

When defining permutation tests, Pesarin (2001) relaxes the null assumption that $\boldsymbol{z}$ is an observation due to i.i.d. continuous random variables, into the assumption that $\boldsymbol{z}$ constitutes a random sample due to an n–dimensional sampling variable $\boldsymbol{Z}$, whose Z_i components are exchangeable with respect to the two groups (hence, Z_i's are not necessarily i.i.d.), and whose population distribution under the null is $F \in \mathfrak{F}$, meaning that F is a member of some nonparametric family $\mathfrak{F}$. In this case $T(\boldsymbol{Z})$ is still a sufficient statistic for F and, if the Z_i's are also continuous random variables, the conditional distribution $\boldsymbol{Z}$ given $T(\boldsymbol{Z})$ is also (1.4). This is a more general definition of permutation tests, due to the assumption of exchangeability.

The definitions given so far reveal the main difference between the randomization and permutation tests – the justification for permuting the data. Permuting the data

in a permutation test is ultimately justified by the following assumptions under the null:

1. the existence of some parent distribution underlying the data;

2. the data are due to random sampling;

3. the observed sample is a result of i.i.d. or exchangeable random variables.

These assumptions are necessary for the set of order statistics, $T(\boldsymbol{Z})$, to be sufficient for F under the null. Furthermore, the sufficiency of $T(\boldsymbol{Z})$ is necessary to construct the distribution of $\boldsymbol{Z}$ under the null (by permuting the data in all $n!$ ways), which in turn is necessary for the permutation distribution of the test statistic. The reasons for permuting the data in a randomization test are more basic:

1. the randomization procedure, $\boldsymbol{T}$, used;

2. the assumption of no group differences under H_0.

We note that no distributional or random sampling assumptions about the observed responses are needed in order to permute the data in a randomization test. In the case of randomization tests, the permutations are with respect to the random assignment $\boldsymbol{T}$, and they stand for all possible ways the observed values could have been associated with $\boldsymbol{T}$. In the case of permutation tests, the data permutations signify all possible values $\boldsymbol{Z}$, or a function of $\boldsymbol{Z}$, could take under the null, given what has already been observed. The assumption of random sampling is very helpful in permutation tests, since it enables easy generalizations about a population. This is not the case for randomization tests. Conclusions from randomization tests directly pertain only to the subjects used in the experiment, and any generalization to other subjects must be made subjectively, based on a similarity argument. We also note that a series of two–sample nonparametric linear rank tests that can be generalized as permutation tests

(e.g., the Wilcoxon rank–sum test and Pitman's permutation test) will result in the same p–value as the randomization test under complete randomization using the same score function (e.g., the simple scores for the Wilcoxon rank–sum test and the actual data values for Pitman's permutation test). In this case, the null distribution under the nonparameric test is identical to the one under the randomization test, despite the difference in the ways we arrive at them. These distributions are identical, however, only when randomization is complete. Despite these differences, both randomization and permutation tests are distribution free, in the sense that no assumption about the type of distribution underlying the data under the null is necessary. Moreover, they both are nonparametric tests, in the sense that neither test specifies a parametric model in the null hypothesis.

Although the idea behind constructing a randomization test is very simple, obtaining the p–values for such tests is not trivial. Approximating randomization tests following Efron's BCD(p) is the subject of this dissertation. Rosenberger and Lachin (2002) distinguish among three techniques that can be used to compute randomization tests: exact, Monte Carlo and asymptotic. Mehta, Patel and Wei (1988a) provided an efficient recursive algorithm for constructing exact conditional randomization tests (tests that use the conditional reference set Ω_c) following restricted randomization, and applied it for $n = 30$. In Mehta *et al.* (1988b), the efficiency of the exact algorithm is described as being quite sensitive to the functional form of the scores of the linear rank statistics and consuming considerable amount of CPU time for moderate–sized data sets. Although their 1988a paper ended with a statement that the algorithm would be integrated into popular software packages, according to a recent communication with Dr. Cyrus Mehta the algorithm has not been implemented in StatXact® or any other package.

The network algorithm by Mehta *et al.* (1988a) is essentially a special case of the

recursive algorithm by Hollander and Peña (1988) to determine the exact distribution of conditional randomization tests following restricted randomization and comparing two or more groups. This algorithm was applied to a data set with a sample of size $n = 37$ and two groups.

It is expected that these exact computational techniques would be able to solve much larger problems with today's computing resources. However, Monte Carlo techniques seem more reasonable for large problems. Mehta, Patel, and Senchaudhuri (1988b) use importance sampling to estimate the conditional randomization test p–value; their technique employs an elegant, but complex, networking algorithm. It can be used with randomization tests based on restricted randomization designs, including Efron's BCD. This technique samples N_w collections, i.e. randomization sequences, from the conditional reference set in proportion to their importance for reducing the variance of the p–value estimator. Let $\boldsymbol{t}_1, ..., \boldsymbol{t}_{N_w}$ be sampled independently and with replacement from Ω_c, each with respective probability $g(\boldsymbol{t}_j)$. The jth sampled collection yields a realization w_j of the random variable

$$W_j = \begin{cases} h(\boldsymbol{t}_j)/g(\boldsymbol{t}_j), & \text{if} \quad \boldsymbol{t}_j \in \Omega^*, \\ 0, & \text{otherwise}, \end{cases}$$

where $h(\boldsymbol{t}_j)$ is the probability of observing $\boldsymbol{t}_j$ given Ω_c and Ω^* is the critical region. The estimator $\overline{W} = \sum_{j=1}^{N_w} W_j/N_w$ is an unbiased estimator of the test p–value. To make the sample variance $\sum_{j=1}^{N_w}(W_j - \overline{W})^2/(N_w - 1)$ as small as possible, g would have to be as close as possible to the ideal

$$g(\boldsymbol{t}_j)_{\text{ideal}} = \begin{cases} h(\boldsymbol{t}_j)/p\text{–value}, & \text{if} \quad \boldsymbol{t}_j \in \Omega^*, \\ 0, & \text{otherwise}, \end{cases}$$

the probability of $\boldsymbol{t}_j$ given Ω^*. Because the p–value is unknown, Mehta *et al.* (1988b) approximate it via asymptotic normality. Thus, each $\boldsymbol{t}_j$ is sampled with probability

$$g(\boldsymbol{t}_j) = \begin{cases} h(\boldsymbol{t}_j)/p(\boldsymbol{t}_j), & \text{if} \quad \boldsymbol{t}_j \in \Omega^*, \\ 0, & \text{otherwise,} \end{cases}$$

where $p(\boldsymbol{t}_j)$ is a large–sample approximation of the p–value. The values of $h(\boldsymbol{t}_j)$ and $p(\boldsymbol{t}_j)$ are obtained sequentially by passing through a network. As a consequence, the Monte Carlo sampling is limited to the rejection set $\Omega^* \subset \Omega_c$, which reduces the variance of the estimator, and each sequence is sampled with a probability higher than the actual probability assigned to that sequence by the randomization procedure. Importance sampling was applied in approximating conditional logrank tests under BCD and GBCD for data from a prostatic–cancer clinical trial with $n = 88$. The efficiency of the estimator in important sampling relies on the convergence to normality of the test statistic, which is doubtful under the BCD(p), as will be referenced shortly. In addition, presumably the importance sampling method has not been extended to stratified data and may fail for certain very large data sets (Cytel Inc., 2007, p. 512).

While authors (e.g., Smythe, 1988) have determined the asymptotic normality of conditional randomization tests under various score functions and randomization procedures, including Smith's (1984) design, Efron's biased coin induces a stationary distribution, and hence the test statistic will not be asymptotically normal. This phenomenon was noted in a number of papers, first by counterexample in Smythe and Wei (1983) for the unconditional test, and then by simulation by Hollander and Peña (1988) for the conditional test.

1.2 Outline of the thesis

We examine randomization–based inference following Efron's biased coin design by taking the Monte Carlo approach. We demonstrate a technique that generates sequences directly from Ω_c, the conditional distribution. In particular, we describe a naive approach that can be implemented very easily in balanced clinical trials, but does not work for imbalanced ones. Then we demonstrate a technique that generates sequences directly from the conditional distribution. The latter technique is unaffected by either imbalance or choice of scores and is simple to implement. In so doing, we find the exact conditional distribution of $N_1(n)$ given $N_1(j),\ 1 \leq j < n$, under the BCD(p), using combinatoric techniques first applied in Markaryan and Rosenberger (2010). Such a technique uses an estimator that is both unbiased and strongly consistent for the randomization test p–value, and reduces the computational burden to a reasonable time for even large clinical trials. Moreover, our technique has a natural extension to conditional stratified randomization tests and randomization tests when sequential monitoring is implemented during the course of the trial.

This dissertation is organized as follows. Chapter 2 describes a method to approximate the conditional test under the BCD by sampling from the unconditional reference set. It introduces the main result by deriving a complex formula (easily programmed) that allows us to directly sample from the conditional distribution. We apply this result to evaluate the type I error rate and power under a location–shift alternative in an example. In Chapter 3 we extend the main result to conditional stratified randomization tests. Chapter 4 presents the sampling framework for conditional randomization tests following complete randomization and BCD(p) when sequential monitoring is implemented during the course of the trial. We draw conclusions in Chapter 5.

Chapter 2: Monte Carlo Methods for Conditional Randomization Tests

This chapter describes two Monte Carlo methods for approximating conditional randomization tests following Efron's BCD(p) when two groups are compared and no adjustment for covariates is made. Generally, the p–value of a conditional randomization test can be approximated by sampling with replacement a sufficiently large number of randomization sequences from either the unconditional or conditional reference set. The first two sections address both cases in detail. The Monte Carlo methods are compared to the exact test for small sample sizes and to each other. We also show how to approximate the power of a randomization test under a shift model.

2.1 Sampling from the unconditional reference set

We start the development of approximating a conditional randomization test by first addressing the computation of an unconditional randomization test by Monte Carlo methods (Zhang and Rosenberger, 2010). For this purpose, a sufficiently large number of sequences, N_u, must be generated from the appropriate unconditional reference set, Ω_u, using the appropriate restricted randomization procedure. Let N_u randomization sequences, $\boldsymbol{T}_1, ..., \boldsymbol{T}_{N_u}$, be sampled independently and with replacement from Ω_u. For an upper–tailed test, the kth sampled sequence induces a Bernoulli random variable

$$U_k = \begin{cases} 1, & \text{if} \quad S(\boldsymbol{T}_k) \geq S_{obs}, \\ 0, & \text{otherwise}, \end{cases}$$

where $S(\boldsymbol{T}_k)$ is the linear rank test statistic as a function of the randomization sequence and S_{obs} is the observed value of $S(\boldsymbol{T}_k)$. The linear rank statistic in the context of randomization–based inference is defined as

$$S(\boldsymbol{T}_k) = \sum_{i=1}^{n}(a_i - \overline{a}_n)T_{ki} = \boldsymbol{a}_n'\boldsymbol{T}_k,$$

with $\boldsymbol{a}_n'$ being the vector of centered scores. The random variable U_k is Bernoulli with parameter $P(S(\boldsymbol{T}_k) \geq S_{obs}) = p_u$, the p–value of the unconditional test. The Monte Carlo estimator of the test p–value is the unbiased estimator $\overline{U} = \sum_{k=1}^{N_u} U_k/N_u$. The requisite number of Monte Carlo samples can be obtained from the constraint $\text{MSE}(\overline{U}) = p_u(1-p_u)/N_u \leq 1/4N_u \leq \epsilon$. For $\epsilon = 0.0001$, $N_u \geq 2500$.

A similar approach may be taken to approximating conditional randomization tests. This time $K > N_u$ randomization sequences must be sampled, $\boldsymbol{T}_1, ..., \boldsymbol{T}_K$, independently and with replacement from Ω_u. This number of Monte Carlo sequences must be large enough such that at least N_u number of sequences satisfy the condition $N_1(n) = n_1$; that is, the condition that the number of subjects allocated to treatment 1 is n_1 when a total of n subjects are being randomized. The requisite number of sequences, K, is a random variable, best described as a negative binomial random variable with parameters $\pi = P(N_1(n) = n_1)$ and $r = N_u$ (Zhang and Rosenberger, 2010). A sequence $\boldsymbol{T}_k = \boldsymbol{t}$ is sampled from Ω_u with probability

$$f(\boldsymbol{t}) = (1/2)\prod_{j=1}^{n-1}(\phi_{j+1})^{t_{j+1}}(1-\phi_{j+1})^{1-t_{j+1}}, \tag{2.1}$$

where t_{j+1} is the observed value of T_{j+1}, 0 or 1, and

$$\phi_{j+1} = \begin{cases} P(T_{j+1} = 1|D_j), & 1 \leq j \leq n-1, \\ 1/2, & j = 0. \end{cases} \tag{2.2}$$

Let N be in the range of K. The jth sampled sequence induces two Bernoulli random variables

$$A_j = \begin{cases} 1, & \text{if} \quad N_1(n) = n_1, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$B_j = \begin{cases} 1, & \text{if} \quad N_1(n) = n_1 \quad \text{and} \quad S(\boldsymbol{T}_j) \geq S_{obs}, \\ 0, & \text{otherwise.} \end{cases}$$

An estimator of the conditional p–value can be formed as $\sum_{j=1}^{N} B_j / \sum_{j=1}^{N} A_j$.

LEMMA 2.1. *The estimator $\sum_{j=1}^{N} B_j / \sum_{j=1}^{N} A_j$ is a strongly consistent estimator of the conditional randomization test p–value.*

PROOF. The random variable A_j is Bernoulli with parameter $P(N_1(n) = n_1)$, and B_j is Bernoulli with the probability parameter $P(N_1(n) = n_1)\, p_c$, since

$$\begin{aligned} P(\{N_1(n) = n_1\} \cap \{S(\boldsymbol{T}_j) \geq S_{obs}\}) &= P(N_1(n) = n_1)\, P(S(\boldsymbol{T}_j) \geq S_{obs} | N_1(n) = n_1) \\ &= P(N_1(n) = n_1) \cdot p_c, \end{aligned}$$

where p_c is the conditional randomization test p-value. By the SLLN, $\sum_{j=1}^{N} A_j / N$ converges a.s. to $P(N_1(n) = n_1)$, and $\sum_{j=1}^{N} B_j / N$ converges a.s. to $P(N_1(n) = n_1)\, p_c$. It follows that the estimator $\sum_{j=1}^{N} B_j / \sum_{j=1}^{N} A_j$ converges a.s. to p_c. □

The expectation and variance of K are

$$E(K) = \frac{r}{\pi} = \frac{N_u}{P(N_1(n) = n_1)}$$

$$\mathrm{Var}(K) = \frac{N_u\{1 - P(N_1(n) = n_1)\}}{P(N_1(n) = n_1)^2} \sim \frac{N_u}{P(N_1(n) = n_1)^2}.$$

The latter approximation applies only when the denominator in the expectation is very small. The number of sequences N_u can be replaced by the lower bound $1/4\epsilon$. The value of the probability $P(N_1(n) = n_1)$ depends on the randomization procedure used. For complete randomization, $P(N_1(n) = n_1) = \binom{n}{n_1}/2^n$. For Efron's BCD($p$), Markaryan and Rosenberger (2010) provide

$$P(N_1(n) = n_1) = \begin{cases} p^{n_1} \displaystyle\sum_{l=0}^{n_1-1} \frac{n-2l}{n+2l}\binom{n_1+l}{l} q^l, & \text{when } n_1 = \frac{n}{2}, \\ \frac{p^{n_1}}{2} \displaystyle\sum_{l=0}^{n_1} \frac{n-n_1-l}{n-n_1+l}\binom{n-n_1+l}{l} q^{n-2n_1+l-1}, & \text{when } 0 \le n_1 < \frac{n}{2}, \\ \frac{p^{n-n_1}}{2} \displaystyle\sum_{l=0}^{n-n_1} \frac{n_1-l}{n_1+l}\binom{n_1+l}{l} q^{2n_1-n+l-1}, & \text{when } \frac{n}{2} < n_1 \le n. \end{cases}$$

With this sampling scheme, $E(K)$ number of Monte Carlo samples for the conditional tests will only be the sufficient number of sequences on the average. Instead, the 95th percentile of K can be used. Table 2.1 reports the Monte Carlo sample sizes for complete and BCD(2/3) randomization procedures when sampling from the unconditional reference set. These sample sizes are reasonable when there is perfect balance in the assignments, but go up considerably in the presence of imbalance. This is because under restricted randomization $P(N_1(n) = n_1)$ is much smaller if $n_1 \neq n/2$ than if $n_1 = n/2$. The smaller $P(N_1(n) = n_1)$, the larger the 95th percentile of K.

Tables 2.2 and 2.3 contain estimates of the upper 0.1 tail probability of the conditional randomization distribution of $\sum_{i=1}^n iT_i$ under various randomization procedures. (Note that, under restricted randomization, the distribution of $\sum_{i=1}^n iT_i$ is not the same as that of $\sum_{i=1}^n r_iT_i$, where r_i is the simple rank of the response on the ith

Table 2.1: *Approximate 95th percentile of K for various n, n_1, $\epsilon = 0.0001$*

n	$n_1 = 0.45n$	$n_1 = 0.48n$	$n_1 = 0.50n$
	$BCD(p = 2/3)$		
100	3,531,344	55,060	5,117
200	3,611,280,266	881,557	5,117
500	$3,877,310 \times 10^{12}$	3,611,026,232	5,117
	$BCD(p = 3/4)$		
100	114,384,212	156,865	3,822
200	$6,754,269 \times 10^{6}$	12,709,307	3,822
500	$1,390,644 \times 10^{21}$	$6,754,269 \times 10^{6}$	3,822
	Complete Randomization		
100	53,240	35,085	32,409
200	124,141	53,707	45,794
500	881,469	107,922	72,377

Table 2.2: *Approximations for the upper* 0.1 *tail of the conditional randomization distribution of the statistic* $\sum_{i=1}^{n} iT_i$ *by sampling from* Ω_u*;* $N_u = 2500$*,* $BCD(0.6)$*.* CPU time in parentheses.

	n_1	Exact	1 MC run	1000 MC runs; mean (S.D.)
$P(\sum_{i=1}^{30} iT_i \geq 254)$	15	0.1057	0.1147 (0.01 min)	0.1056 (0.0061)
$P(\sum_{i=1}^{30} iT_i \geq 209)$	12	0.1009	0.0936 (0.1 min)	0.1010 (0.0061)
$P(\sum_{i=1}^{40} iT_i \geq 441)$	20	0.1011	0.0968 (0.01 min)	0.1010 (0.0061)
$P(\sum_{i=1}^{40} iT_i \geq 362)$	16	0.1000	0.1059 (0.3 min)	0.1001 (0.0060)
$P(\sum_{i=1}^{100} iT_i \geq 2,607)$	50		0.1029 (0.03 min)	0.1055 (0.0060)
$P(\sum_{i=1}^{100} iT_i \geq 2,133)$	40		0.1000 (2 h)	0.1016 (0.0056)
$P(\sum_{i=1}^{500} iT_i \geq 62,924)$	250		0.1020 (0.2 min)	0.1100 (0.0061)
$P(\sum_{i=1}^{500} iT_i \geq 51,100)$	200		impractical*	

* A Monte Carlo sample size of $4,255,535 \times 10^{15}$ is required.

Table 2.3: *Approximations for the upper* 0.1 *tail of the conditional randomization distribution of the statistic* $\sum_{i=1}^{n} iT_i$ *by sampling from* Ω_u*;* $N_u = 2500$*, complete randomization. CPU time in parentheses.*

	n_1	Exact	1 MC run	1000 MC runs; mean (S.D.)
$P(\sum_{i=1}^{30} iT_i \geq 264)$	15	0.1008	0.1147 (0.01 min)	0.1010 (0.0060)
$P(\sum_{i=1}^{30} iT_i \geq 217)$	12	0.1000	0.1013 (0.02 min)	0.0996 (0.0060)
$P(\sum_{i=1}^{40} iT_i \geq 458)$	20	0.1006	0.0963 (0.02 min)	0.1008 (0.0061)
$P(\sum_{i=1}^{40} iT_i \geq 375)$	16	0.1009	0.0971 (0.03 min)	0.1009 (0.0061)
$P(\sum_{i=1}^{100} iT_i \geq 2,707)$	50		0.1101 (0.06 min)	0.1061 (0.0059)
$P(\sum_{i=1}^{100} iT_i \geq 2,199)$	40		0.1078 (0.4 min)	0.1049 (0.0061)
$P(\sum_{i=1}^{500} iT_i \geq 64,682)$	250		0.0991 (0.6 min)	0.1015 (0.0060)
$P(\sum_{i=1}^{500} iT_i \geq 52,137)$	200		impractical*	

* A Monte Carlo sample size of $1,672,526,419$ is required.

assignment. They are, however, the same under complete randomization. The statistic $\sum_{i=1}^{n} iT_i$ would be encountered in data with strong time trends.) The CPU times reported for one Monte Carlo run are for SAS on a Dell OPTIPLEX 745 desktop with an Intel Core 2 Duo processor. The Monte Carlo estimated probabilities are very close to the exact ones, for sample sizes allowing the computation of such probabilities. The variability in the estimates is roughly the same across all sample sizes. Computationally, this sampling technique works very well when the assignments in each group are equal or almost equal. Moreover, it will work well with balanced and almost balanced designs. The technique is inefficient and impractical when there is a large imbalance in the observed assignments and the sample size n is large, in particular if such an imbalance occurs under BCD with p close to 1. A few situations of this type appear in Tables 2.2 and 2.3 and were marked as "impractical". The

inefficiency is due to the enormous number of sequences that must be sampled from Ω_u in presence of noticeable imbalance.

2.2 Sampling from the conditional reference set

Rather than sampling too many sequences and discarding those that do not satisfy the condition $N_1(n) = n_1$ as was done previously, it is more efficient to sample directly from Ω_c – the collection of all randomization sequences that satisfy the condition $N_1(n) = n_1$. This is the idea at the center of this second method, which is a major contribution of this chapter. The set Ω_c will be called the conditional randomization set. Let N_c randomization sequences, $\boldsymbol{T}_1, ..., \boldsymbol{T}_{N_c}$, be sampled independently and with replacement strictly from Ω_c, each with respective probabilities $h(\boldsymbol{t}_1), ..., h(\boldsymbol{t}_{N_c})$. For an upper–tailed test, the kth sampled sequence induces a Bernoulli random variable

$$R_k = \begin{cases} 1, & \text{if} \quad S(\boldsymbol{T}_k) \geq S_{obs}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

The Monte Carlo estimator of the two–sided test p–value is the unbiased estimator $\overline{R} = \sum_{k=1}^{N_c} R_k/N_c$. As in the case of unconditional tests, the requisite number of Monte Carlo samples can be obtained from the constraint $\text{MSE}(\overline{R}) = p_c(1-p_c)/N_c \leq 1/4N_c \leq \epsilon$. For $\epsilon = 0.0001$, $N_c \geq 2500$. Higher precision in estimation is possible by finding N_c that ensures $P(|\overline{R} - p_c| \leq 0.1p_c) = 0.99$, for instance. Let $\mathcal{Z}$ be a standard normal random variable. Using the Central Limit Theorem,

$$P(|\overline{R} - p_c| \leq 0.1p_c) = P\left(\frac{|\overline{R} - p_c|}{\sqrt{\frac{p_c(1-p_c)}{N_c}}} \leq \frac{0.1p_c}{\sqrt{\frac{p_c(1-p_c)}{N_c}}}\right) \approx P\left(|\mathcal{Z}| \leq \frac{0.1p_c}{\sqrt{\frac{p_c(1-p_c)}{N_c}}}\right) = 0.99.$$

It follows that $N_c \approx (2.576/0.1)^2(1-p_c)/p_c$. Thus, to estimate a p–value as large as 0.04 with an error of 10% of 0.04 with 0.99 probability, the Monte Carlo sample size must be $N_c = 15,924$. If a smaller p–value is expected, N_c will be larger.

LEMMA 2.2. *The estimator $\overline{R}$ is a strongly consistent estimator of the conditional randomization test p–value.*

PROOF. Let $\boldsymbol{T}_1, ..., \boldsymbol{T}_{N_c}$ be sequences of n assignments sampled independently and with replacement from Ω_c. For an upper–tailed conditional randomization test

$$P\left(S(\boldsymbol{T}_j) \geq S_{obs}\right) = P\left(S(\boldsymbol{T}_j) \geq S_{obs} | N_1(n) = n_1\right) = p_c,$$

because $\boldsymbol{T}_j$ is guaranteed to satisfy $N_1(n) = n_1$. It follows that R_j is distributed as a Bernoulli random variable with the probability parameter p_c. For this method, the Monte Carlo estimator of p_c is the unbiased estimator $\overline{R} = \sum_{j=1}^{N_c} R_j/N_c$. By the SLLN, $\overline{R}$ converges a.s. to p_c. □

Next, we address the sampling probabilities $h(\boldsymbol{t}_i), ..., h(\boldsymbol{t}_{N_c})$. The allocation rule for the BCD(p) and complete randomization can both be expressed more generally as

$$\phi_{j+1}(m_j) = \begin{cases} P(T_{j+1} = 1 | N_1(j) = m_j), & 1 \leq j \leq n-1, \\ 1/2, & j = 0, \end{cases} \tag{2.4}$$

where $0 \leq m_j \leq j$. (Conditioning on $\{N_1(j) = m_j\}$ is equivalent to conditioning on $\{D_j = 2m_j - j\}$.) This rule (used n times) will yield a random sequence that will not necessarily be in Ω_c. To guarantee a sequence from Ω_c, the conditioning set in (2.4) must be updated from $\{N_1(j) = m_j\}$ to $\{N_1(j) = m_j, N_1(n) = n_1\}$. Consequently, the rule

$$p_{j+1} = \begin{cases} P(T_{j+1} = 1 | N_1(j) = m_j, N_1(n) = n_1), & 1 \leq j \leq n-1, \\ P(T_1 = 1 | N_1(n) = n_1), & j = 0. \end{cases} \tag{2.5}$$

guarantees a random sequence strictly from Ω_c. The procedure in (2.5) is expressed in terms of $\phi_{j+1}(m_j)$ as follows:

$$\begin{aligned} p_{j+1} &= P\left(T_{j+1} = 1 | N_1(j) = m_j, N_1(n) = n_1\right) \\ &= \frac{P\left(N_1(n) = n_1, T_{j+1} = 1, N_1(j) = m_j\right)}{P\left(N_1(j) = m_j, N_1(n) = n_1\right)} \\ &= \frac{P\left(N_1(n) = n_1 | T_{j+1} = 1, N_1(j) = m_j\right) P\left(T_{j+1} = 1 | N_1(j) = m_j\right)}{P\left(N_1(n) = n_1 | N_1(j) = m_j\right) P\left(N_1(j) = m_j\right)} \\ &\qquad \times P\left(N_1(j) = m_j\right) \\ &= \begin{cases} \phi_{j+1}(m_j) \dfrac{P\left(N_1(n) = n_1 | T_{j+1} = 1, N_1(j) = m_j\right)}{P\left(N_1(n) = n_1 | N_1(j) = m_j\right)}, & 1 \leq j \leq n-1, \\ 1/2 \dfrac{P\left(N_1(n) = n_1 | T_{j+1} = 1\right)}{P\left(N_1(n) = n_1\right)}, & j = 0. \end{cases} \end{aligned} \tag{2.6}$$

Furthermore, for $k = 1, ..., N_c$, a sequence $\boldsymbol{T}_k = \boldsymbol{t}$ is sampled from Ω_c with probability

$$h(\boldsymbol{t}) = \prod_{j=0}^{n-1} (p_{j+1})^{t_{j+1}} (1 - p_{j+1})^{1 - t_{j+1}}. \tag{2.7}$$

where t_{j+1} is the observed value of T_{j+1}, 0 or 1. Before we discuss the evaluation of p_{j+1} under the BCD(p), which is the main result of this section, we address the Markovian property of $N_1(j)$ for restricted randomization procedures defined by (2.4).

LEMMA 2.3. *Let* $\phi_{j+1} = P(T_{j+1} = 1|D_j)$ *and let* $N_1(j) = \sum_{i=1}^{j} T_i$. *The sequence of random variables* $\{N_1(j) = m_j\}_{j\geq 1}$ *is a Markov chain of order 1, where* $0 \leq m_j \leq j$.

PROOF. Because $\phi_{j+1} = P(T_{j+1} = 1|D_j = 2m_j - j) = P(T_{j+1} = 1|N_1(j) = m_j)$,

$$\begin{aligned} &P\left(N_1(j+1) = m_{j+1}|N_1(j) = m_j, ..., N_1(1) = m_1\right) \\ &\qquad = P\left(T_{j+1} = m_{j+1} - m_j|N_1(j) = m_j, ..., N_1(1) = m_1\right) \\ &\qquad = P\left(T_{j+1} = m_{j+1} - m_j|D_j = 2m_j - j\right) \\ &\qquad = P\left(T_{j+1} = m_{j+1} - m_j|N_1(j) = m_j\right) \\ &\qquad = P\left(N_1(j+1) = m_{j+1}|N_1(j) = m_j\right), \end{aligned}$$

where the second equality follows from the definition of ϕ_{j+1}. □

Let q be a positive integer and $C \subset \{N_1(j) = m_j\}$. A consequence of the Markovian property is $P\left(N_1(j+q) = m_{j+q}|C\right) = P\left(N_1(j+q) = m_{j+q}|N_1(j) = m_j\right)$. Using Lemma 2.3, we note that p_{j+1} as defined in (2.6) is equivalent to

$$p_{j+1} = \begin{cases} \phi_{j+1}(m_j)\dfrac{P\left(N_1(n) = n_1|N_1(j+1) = m_j + 1\right)}{P\left(N_1(n) = n_1|N_1(j) = m_j\right)}, \ 1 \leq j \leq n-1, \\ \\ 1/2\dfrac{P\left(N_1(n) = n_1|T_{j+1} = 1\right)}{P\left(N_1(n) = n_1\right)}, \ j = 0, \end{cases} \tag{2.8}$$

since $C = \{T_{j+1} = 1, N_1(j) = m_j\} \subset \{N_1(j+1) = m_j + 1\}$.

The conditional probabilities in p_{j+1} are largely intractable for most restricted randomization procedures, for example, the generalized biased coin designs of Smith (1984). For complete randomization, p_{j+1} is given by the following lemma.

LEMMA 2.4. *For complete randomization* $p_{j+1} = (n_1 - m_j)/(n - j),\ 0 \le j \le n-1$ *and* $m_0 = 0$.

PROOF. We have

$$p_{j+1} = (1/2)\frac{\binom{n-j-1}{n_1-m_j-1}(1/2)^{n-j-1}}{\binom{n-j}{n_1-m_j}(1/2)^{n-j}} = \frac{n_1-m_j}{n-j},\ 0 \le j \le n-1 \text{ and } m_0 = 0. \ \square$$

This result is the random allocation rule when $n_1 = n/2$ (Rosenberger and Lachin, 2002), which is sometimes used to fill permuted blocks. For the BCD(p), the closed form solution for $P(N_1(n) = n_1 | N_1(j) = m_j)$ is given in Theorem 2.1., which must be applied to the numerator and denominator of (2.8) to evaluate p_{j+1} at each step j. The sampling procedure then follows by simply generating N_c sequences from Ω_c using (2.8) sequentially. In practice, $N_c = 2500$ yields very accurate results.

THEOREM 2.1. *For the BCD(p):*

1. *When* $1 \le j < n$ *and* $0 \le m_j < j/2$, $P(N_1(n) = n_1 | N_1(j) = m_j)$ *is*

$$\begin{cases} \binom{n-j}{n_1-m_j} p^{n_1-m_j} q^{n-j-n_1+m_j}, & m_j \le n_1 < j - m_j; \\ 0.5p^{n_1-m_j} \displaystyle\sum_{l=0}^{n_1+m_j-j} \frac{n-n_1-m_j-l}{n-n_1-m_j+l}\binom{n-n_1-m_j+l}{l} q^{n-2n_1-1+l} \\ \quad + p^{n_1-m_j} q^{n-j-n_1+m_j}\left(\binom{n-j}{n_1-m_j} - \binom{n-j}{n_1-j+m_j}\right), & j - m_j \le n_1 < n/2; \\ p^{n_1-m_j} \displaystyle\sum_{l=0}^{n-j-n_1+m_j} \frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l} q^l, & n_1 = n/2; \\ 0.5p^{n-n_1-m_j} \displaystyle\sum_{l=0}^{n-j-n_1+m_j} \frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l} q^{2n_1-n-1+l}, & \\ & n/2 < n_1 \le n-j+m_j. \end{cases}$$

2. *When* $0 \leq j < n$ *and* $m_j = j/2$,

$$P(N_1(n) = n_1 | N_1(j) = m_j) = P(N_1(n-j) = n_1 - m_j), \quad m_j \leq n_1 \leq n - j + m_j,$$

where the unconditional distribution is derived in Markaryan and Rosenberger (2010).

3. *When* $1 \leq j < n$ *and* $j/2 < m_j \leq j$, $P(N_1(n) = n_1 | N_1(j) = m_j)$ *is*

$$\begin{cases} 0.5p^{n_1+m_j-j} \displaystyle\sum_{l=0}^{n_1-m_j} \frac{n-j-n_1+m_j-l}{n-j-n_1+m_j+l} \binom{n-j-n_1+m_j+l}{l} q^{n-2n_1-1+l}, & \\ & m_j \leq n_1 < n/2; \\ p^{n-j-n_1+m_j} \displaystyle\sum_{l=0}^{n_1-m_j} \frac{n-j-n_1+m_j-l}{n-j-n_1+m_j+l} \binom{n-j-n_1+m_j+l}{l} q^{l}, & \\ & n_1 = n/2; \\ 0.5p^{n-j-n_1+m_j} \displaystyle\sum_{l=0}^{n-n_1-m_j} \frac{n_1+m_j-j-l}{n_1+m_j-j+l} \binom{n_1+m_j-j+l}{l} q^{2n_1-n-1+l} & \\ \quad + p^{n-j-n_1+m_j} q^{n_1-m_j} \left(\binom{n-j}{n_1-m_j} - \binom{n-j}{n_1-j+m_j} \right), & \\ & n/2 < n_1 \leq n - m_j; \\ \binom{n-j}{n_1-m_j} p^{n-j-n_1+m_j} q^{n_1-m_j}, & n - m_j < n_1 \leq n - j + m_j. \end{cases}$$

PROOF. The case $0 \leq m_j < j/2$ is proved by induction. The proof is similar when $j/2 < m_j \leq j$, due to symmetry. As in Markaryan and Rosenberger (2010), the combinatorial identity

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \tag{2.9}$$

is used several times in this proof. Let n, j and m_j be integers such that $1 \leq j < n$ and $0 \leq m_j < j/2$. The theorem holds for $n = 2$ and $n = 3$. Assume that when $0 \leq m_j < j/2$ the theorem is true for some positive integer n. It must be proved that

the theorem holds for $n+1$. If n is odd, the proof has five cases: $m_j \leq n_1 < j - m_j$, $n_1 = j - m_j$, $j - m_j < n_1 < (n+1)/2$, $n_1 = (n+1)/2$, and $(n+1)/2 < n_1 \leq n+1-j+m_j$. If n is even, the following cases must be considered: $m_j \leq n_1 < j - m_j$, $n_1 = j - m_j$ and $n_1 < n/2$, $n_1 = j - m_j$ and $n_1 = n/2$, $j - m_j < n_1 < n/2$, $j - m_j < n_1$ and $n_1 = n/2$, $n_1 = n/2 + 1$ and $n/2 + 1 < n_1 \leq n + 1 - j + m_j$. The proofs for these cases are either identical or very similar in technique to those when n is odd, and thus are omitted. If n is odd, the proof is as follows.

Case $m_j \leq n_1 < j - m_j$: It must be shown that

$$P(N_1(n+1) = n_1 | N_1(j) = m_j) = \binom{n+1-j}{n_1 - m_j} p^{n_1 - m_j} q^{n+1-j-n_1+m_j}. \tag{2.10}$$

$$\begin{aligned}
&P(N_1(n+1) = n_1 | N_1(j) = m_j) \\
&= pP(N_1(n) = n_1 - 1 | N_1(j) = m_j) + qP(N_1(n) = n_1 | N_1(j) = m_j) \\
&= p\binom{n-j}{n_1 - 1 - m_j} p^{n_1 - 1 - m_j} q^{n-j-n_1+1+m_j} + q\binom{n-j}{n_1 - m_j} p^{n_1 - m_j} q^{n-j-n_1+m_j} \\
&= \left(\binom{n-j}{n_1 - 1 - m_j} + \binom{n-j}{n_1 - m_j} \right) p^{n_1 - m_j} q^{n+1-j-n_1+m_j} \\
&= \binom{n+1-j}{n_1 - m_j} p^{n_1 - m_j} q^{n+1-j-n_1+m_j},
\end{aligned}$$

where the last equality is due to the combinatorial identity in (2.9).

Case $n_1 = j - m_j$: It must be shown that $P(N_1(n+1) = j - m_j | N_1(j) = m_j)$

$$= \frac{1}{2} p^{j-2m_j} q^{n-2(j-m_j)} + p^{j-2m_j} q^{n+1-2(j-m_j)} \left(\binom{n+1-j}{j-2m_j} - 1 \right). \tag{2.11}$$

Then $P(N_1(n+1) = j - m_j | N_1(j) = m_j)$

$$= pP(N_1(n) = j - m_j - 1 | N_1(j) = m_j) + qP(N_1(n) = j - m_j | N_1(j) = m_j)$$

$$= p\binom{n-j}{j-1-2m_j} p^{j-1-2m_j} q^{n+1-2(j-m_j)}$$

$$+ q\frac{1}{2} p^{j-m_j-m_j} \sum_{l=0}^{j-m_j+m_j-j} \frac{n-j+m_j-m_j-l}{n-j+m_j-m_j+l} \binom{n-j+m_j-m_j+l}{l} q^{n-2(j-m_j)-1+l}$$

$$+ qp^{j-m_j-m_j} q^{n-j-j+m_j+m_j} \left(\binom{n-j}{j-m_j-m_j} - \binom{n-j}{j-m_j-j+m_j} \right)$$

$$= \frac{1}{2} p^{j-2m_j} q^{n-2(j-m_j)} + p^{j-2m_j} q^{n+1-2(j-m_j)} \left(\binom{n-j}{j-1-2m_j} + \binom{n-j}{j-2m_j} - 1 \right)$$

$$= \frac{1}{2} p^{j-2m_j} q^{n-2(j-m_j)} + p^{j-2m_j} q^{n+1-2(j-m_j)} \left(\binom{n+1-j}{j-2m_j} - 1 \right),$$

since $j - m_j + m_j - j = 0$ in the sum and after using (2.9) in the last line.

Case $j - m_j < n_1 < (n+1)/2$: It must be shown that

$$P(N_1(n+1) = n_1 | N_1(j) = m_j)$$

$$= \frac{1}{2} p^{n_1-m_j} \sum_{l=0}^{n_1+m_j-j} \frac{n+1-n_1-m_j-l}{n+1-n_1-m_j+l} \binom{n+1-n_1-m_j+l}{l} q^{n-2n_1+l}$$

$$+ p^{n_1-m_j} q^{n+1-j-n_1+m_j} \left(\binom{n+1-j}{n_1-m_j} - \binom{n+1-j}{n_1-j+m_j} \right). \tag{2.12}$$

Similarly, $P(N_1(n+1) = n_1 | N_1(j) = m_j)$

$$= pP(N_1(n) = n_1 - 1 | N_1(j) = m_j) + qP(N_1(n) = n_1 | N_1(j) = m_j)$$

$$= p\frac{1}{2}p^{n_1-1-m_j}\sum_{l=0}^{n_1-1+m_j-j}\frac{n-n_1+1-m_j-l}{n-n_1+1-m_j+l}\binom{n-n_1+1-m_j+l}{l}q^{n-2(n_1-1)-1+l}$$

$$+pp^{n_1-1-m_j}q^{n-j-n_1+1+m_j}\left(\binom{n-j}{n_1-1-m_j}-\binom{n-j}{n_1-1-j+m_j}\right)$$

$$+q\frac{1}{2}p^{n_1-m_j}\sum_{l=0}^{n_1+m_j-j}\frac{n-n_1-m_j-l}{n-n_1-m_j+l}\binom{n-n_1-m_j+l}{l}q^{n-2n_1-1+l}$$

$$+qp^{n_1-m_j}q^{n-j-n_1+m_j}\left(\binom{n-j}{n_1-m_j}-\binom{n-j}{n_1-j+m_j}\right)$$

$$= \frac{1}{2}p^{n_1-m_j}\sum_{l=0}^{n_1-1+m_j-j}\frac{n-n_1+1-m_j-l}{n-n_1+1-m_j+l}\binom{n-n_1+1-m_j+l}{l}q^{n-2n_1+1+l}$$

$$+\frac{1}{2}p^{n_1-m_j}\sum_{l=0}^{n_1+m_j-j}\frac{n-n_1-m_j-l}{n-n_1-m_j+l}\binom{n-n_1-m_j+l}{l}q^{n-2n_1+l}$$

$$+p^{n_1-m_j}q^{n+1-j-n_1+m_j}\left(\binom{n+1-j}{n_1-m_j}-\binom{n+1-j}{n_1-j+m_j}\right),$$

where the last term follows from using (2.9) twice. It remains to show that

$$\sum_{l=0}^{n_1+m_j-j}\frac{n+1-n_1-m_j-l}{n+1-n_1-m_j+l}\binom{n+1-n_1-m_j+l}{l}q^{n-2n_1+l}$$

$$= \sum_{l=0}^{n_1-1+m_j-j}\frac{n-n_1+1-m_j-l}{n-n_1+1-m_j+l}\binom{n-n_1+1-m_j+l}{l}q^{n-2n_1+1+l}$$

$$+\sum_{l=0}^{n_1+m_j-j}\frac{n-n_1-m_j-l}{n-n_1-m_j+l}\binom{n-n_1-m_j+l}{l}q^{n-2n_1+l},$$

which, after substituting $k = l+1$ in the second to last summation, is the same as

showing that

$$\sum_{l=0}^{n_1+m_j-j} \frac{n+1-n_1-m_j-l}{n+1-n_1-m_j+l}\binom{n+1-n_1-m_j+l}{l} q^{n-2n_1+l}$$

$$- \sum_{l=0}^{n_1+m_j-j} \frac{n-n_1-m_j-l}{n-n_1-m_j+l}\binom{n-n_1-m_j+l}{l} q^{n-2n_1+l}$$

$$= \sum_{k=1}^{n_1+m_j-j} \frac{n-n_1-m_j+2-k}{n-n_1-m_j+k}\binom{n-n_1-m_j+k}{k-1} q^{n-2n_1+k}.$$

But this follows from using the identity

$$\frac{v+1-l}{v+1+l}\binom{v+1+l}{l} - \frac{v-l}{v+l}\binom{v+l}{l} = \frac{v+2-l}{v+l}\binom{v+l}{l-1} \tag{2.13}$$

inside the summations with $v = n - n_1 - m_j$. The identity (2.13) is verified below:

$$\frac{v+1-l}{v+1+l}\binom{v+1+l}{l} - \frac{v-l}{v+l}\binom{v+l}{l} = \frac{v+1-l}{v+1}\binom{v+l}{l} - \frac{v-l}{v+l}\binom{v+l}{l}$$

$$= \frac{l(v+2-l)}{(v+1)(v+l)}\binom{v+l}{l}$$

$$= \frac{v+2-l}{v+l}\binom{v+l}{l-1}.$$

The proof of this case is complete.

Case $n_1 = (n+1)/2$: It must be shown that $P(N_1(n+1) = n_1 | N_1(j) = m_j)$

$$= p^{n_1-m_j} \sum_{l=0}^{n+1-j-n_1+m_j} \frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l} q^l. \tag{2.14}$$

In this case, the algebraic manipulations make use of the following facts:

$$n-(n_1-1)=n_1,\ \ n-2(n_1-1)=1,\ \ 2n_1-n=1,\ \ n_1-1+m_j-j=n-j-n_1+m_j.$$

This time,

$$P(N_1(n+1)=n_1|N_1(j)=m_j)$$

$$=pP(N_1(n)=n_1-1|N_1(j)=m_j)+pP(N_1(n)=n_1|N_1(j)=m_j)$$

$$=p\frac{1}{2}p^{n_1-1-m_j}\sum_{l=0}^{n_1-1+m_j-j}\frac{n-(n_1-1)-m_j-l}{n-(n_1-1)-m_j+l}\binom{n-(n_1-1)-m_j+l}{l}q^{n-2(n_1-1)-1+l}$$

$$+pp^{n_1-1-m_j}q^{n-j-(n_1-1)+m_j}\left(\binom{n-j}{n_1-1-m_j}-\binom{n-j}{n_1-1-j+m_j}\right)$$

$$+p\frac{1}{2}p^{n-n_1-m_j}\sum_{l=0}^{n-j-n_1+m_j}\frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l}q^{2n_1-n-1+l}$$

$$=p^{n_1-m_j}\sum_{l=0}^{n-j-n_1+m_j}\frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l}q^l$$

$$+p^{n_1-m_j}q^{n+1-j-n_1+m_j}\left(\binom{n-j}{n_1-1-m_j}-\binom{n-j}{n_1-1-j+m_j}\right).$$

One still has to show that the last term of the preceding sum is the last term in (2.14). This is accomplished by showing that the two corresponding binomial coefficients are identical. The parts involving p and q are identical, and so is the remaining part, after verifying that

$$\binom{n-j}{n_1-1-m_j}-\binom{n-j}{n_1-1-j+m_j}=\frac{j-2m_j}{n+1-j}\binom{n+1-j}{n_1-m_j}.$$

Since $n_1 - 1 + m_j - j = n - j - n_1 + m_j$ (from $n_1 = (n+1)/2$),

$$\begin{aligned}\binom{n-j}{n_1-1-m_j} - \binom{n-j}{n_1-1-j+m_j} &= \binom{n-j}{n_1-1-m_j} - \binom{n-j}{n_1-m_j} \\ &= \frac{n_1-m_j}{n+1-j}\binom{n+1-j}{n_1-m_j} \\ &\quad - \frac{n_1-j+m_j}{n+1-j}\binom{n+1-j}{n_1-m_j} \\ &= \frac{j-2m_j}{n+1-j}\binom{n+1-j}{n_1-m_j}.\end{aligned}$$

Case $(n+1)/2 + 1 < n_1 \leq n + 1 - j + m_j$: It must be shown that

$$P(N_1(n+1) = n_1 | N_1(j) = m_j) =$$

$$\frac{1}{2}p^{n+1-n_1-m_j} \sum_{l=0}^{n+1-j-n_1+m_j} \frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l} q^{2n_1-n-2+l}. \tag{2.15}$$

We see that

$$P(N_1(n+1) = n_1 | N_1(j) = m_j)$$

$$= qP(N_1(n) = n_1 - 1 | N_1(j) = m_j) + pP(N_1(n) = n_1 | N_1(j) = m_j)$$

$$= q\frac{1}{2}p^{n-(n_1-1)-m_j} \sum_{l=0}^{n-j-(n_1-1)+m_j} \frac{n_1-1-m_j-l}{n_1-1-m_j+l}\binom{n_1-1-m_j+l}{l} q^{2(n_1-1)-n-1+l}$$

$$+p\frac{1}{2}p^{n-n_1-m_j} \sum_{l=0}^{n-j-n_1+m_j} \frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l} q^{2n_1-n-1+l}$$

$$= \frac{1}{2}p^{n+1-n_1-m_j} \sum_{l=0}^{n+1-j-n_1+m_j} \frac{n_1-1-m_j-l}{n_1-1-m_j+l}\binom{n_1-1-m_j+l}{l} q^{2n_1-n-2+l}$$

$$+\frac{1}{2}p^{n+1-n_1-m_j} \sum_{l=0}^{n-j-n_1+m_j} \frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l} q^{2n_1-n-2+(l+1)}.$$

After substituting $k = l + 1$ in the last sum, it remains to show that

$$\sum_{l=0}^{n+1-j-n_1+m_j} \frac{n_1-m_j-l}{n_1-m_j+l}\binom{n_1-m_j+l}{l} q^{2n_1-n-2+l}$$

$$- \sum_{l=0}^{n+1-j-n_1+m_j} \frac{n_1-1-m_j-l}{n_1-1-m_j+l}\binom{n_1-1-m_j+l}{l} q^{2n_1-n-2+l}$$

$$= \sum_{k=1}^{n+1-j-n_1+m_j} \frac{n_1+1-m_j-k}{n_1-1-m_j+k}\binom{n_1-1-m_j+k}{k-1} q^{2n_1-n-2+k}.$$

By an application of (2.13) with $v = n_1 - 1 - m_j$, this case is demonstrated.

The case $0 \leq j < n$ and $m_j = j/2$ is proved next. Here one simply has to observe that the conditional distribution $P(N_1(n) = n_1 | N_1(j) = m_j)$ is the probability of observing any sequence of $n - j$ assignments that transforms the pair (j, m_j) into (n, n_1). Since every time $m_j = j/2$, the probability of the next assignment is reset to $1/2$, the sequences that transform (j, m_j) into (n, n_1) are the same sequences that transform $(0, 0)$ into $(n - j, n_1 - m_j)$. This is exactly what the unconditional distribution describes. This completes the proof. □

The Monte Carlo method based on sampling directly from the conditional randomization set works very well for complete randomization and BCD(p) with any value for p. Tables 2.4 and 2.5 provide approximations for the upper 0.1 tail of the Wilcoxon statistic under complete and BCD(0.6) randomization. For small sample

sizes, the approximations are very close to the true probabilities in both cases. As expected, the variability of the estimates does not change across different sample sizes n. The method is very fast in the case of complete randomization, even for samples as large as 1000, because of the simplicity of p_{j+1}. The computational complexity of this sampling scheme for BCD(p) is higher than for complete randomization because of the burden associated with evaluating p_{j+1} for large sample sizes, but is unaffected by the value of p.

For two–sided tests, we note that the kth sampled randomization sequence induces a Bernoulli random variable

$$R_k = \begin{cases} 1, & \text{if} \quad |S(\boldsymbol{T}_k) - \mu_{|n_1}| \geq |S_{obs} - \mu_{|n_1}|, \\ 0, & \text{otherwise}, \end{cases} \tag{2.16}$$

where $\mu_{|n_1} = E(S(\boldsymbol{T}_k)|N_1(n) = n_1) = \boldsymbol{a}_n' E\left(\boldsymbol{T}_k|N_1(n) = n_1\right)$. The following lemma gives a simple computational expression for $E\left(\boldsymbol{T}_k|N_1(n) = n_1\right)$. Theorem 2.1 is used to evaluate this expression.

LEMMA 2.5. *Let* $\phi_i(j) = P(T_{ki} = 1|N_1(i-1) = j)$ *and* $1 \leq i \leq n$. *The* i*th element of* $E\left(\boldsymbol{T}_k|N_1(n) = n_1\right)$ *is given by*

$$E(T_{ki}|N_1(n) = n_1) = \frac{\sum_{j=0}^{i-1} P(N_1(i-1) = j)\phi_i(j)P(N_1(n) = n_1|N_1(i) = j+1)}{P(N_1(n) = n_1)},$$

where $E(T_{k1}|N_1(n) = n_1) = 1/2P(N_1(n) = n_1|N_1(1) = 1)/P(N_1(n) = n_1)$.

PROOF.

$$\begin{aligned} E(T_{ki}|N_1(n) = n_1) &= \frac{P(T_{ki} = 1, N_1(n) = n_1)}{P(N_1(n) = n_1)} \\ &= \frac{\sum_{j=0}^{i-1} P(N_1(i-1) = j, T_{ki} = 1, N_1(n) = n_1)}{P(N_1(n) = n_1)}. \end{aligned}$$

Table 2.4: *Approximations for the upper* 0.1 *tail of the randomization distribution of the statistic* $\sum_{i=1}^{n} iT_i$ *by sampling from* Ω_c*;* $N_c = 2500$*, BCD(*0.6*). CPU time in parentheses.*

	n_1	Exact	1 MC run	1000 MC runs; mean (S.D.)
$P(\sum_{i=1}^{30} iT_i \geq 254)$	15	0.1057	0.1052 (0.1 min)	0.1053 (0.0061)
$P(\sum_{i=1}^{30} iT_i \geq 209)$	12	0.1009	0.1004 (0.1 min)	0.1008 (0.0059)
$P(\sum_{i=1}^{40} iT_i \geq 441)$	20	0.1011	0.1092 (0.2 min)	0.1009 (0.0061)
$P(\sum_{i=1}^{40} iT_i \geq 362)$	16	0.1000	0.0992 (0.2 min)	0.0997 (0.0060)
$P(\sum_{i=1}^{100} iT_i \geq 2,607)$	50		0.1024 (1.2 min)	0.1055 (0.0060)
$P(\sum_{i=1}^{100} iT_i \geq 2,133)$	40		0.1004 (1 min)	0.1043 (0.0062)
$P(\sum_{i=1}^{500} iT_i \geq 62,924)$	250		0.0960 (30 min)	0.1104 (0.0063)
$P(\sum_{i=1}^{500} iT_i \geq 51,100)$	200		0.1020 (20 min)	0.1030 (0.0058)

The last equality is equivalent to

$$\frac{\sum_{j=0}^{i-1} P(N_1(i-1)=j)P(T_{ki}=1|N_1(i-1)=j)P(N_1(n)=n_1|N_1(i)=j+1)}{P(N_1(n)=n_1)}$$

by Bayes Theorem and Lemma 2.3. When $i = 1$, one has to apply Bayes Theorem to $P(T_{k1} = 1, N_1(n) = n_1)/P(N_1(n) = n_1)$ and the result follows. □

We have identified simple to implement Monte Carlo methods for computing conditional randomization tests. Both Monte Carlo sampling techniques are equivalent for estimating the p–value of a conditional randomization test in the sense that both use estimators that converge $a.s.$ to the test p–value. This follows from Lemmas 2.1 and 2.2 where both sampling schemes were shown to use estimators that converges almost surely to p_c. Sampling from the conditional reference set is preferable because it

Table 2.5: *Approximations for the upper* 0.1 *tail of the conditional randomization distribution of the statistic* $\sum_{i=1}^{n} iT_i$ *by sampling from* Ω_c*;* $N_c = 2500$*, complete randomization. CPU time in parentheses.*

	n_1	Exact	1 MC run	1000 MC runs; mean (S.D.)
$P(\sum_{i=1}^{30} iT_i \geq 264)$	15	0.1008	0.0964 (0.01 min)	0.1006 (0.0058)
$P(\sum_{i=1}^{30} iT_i \geq 217)$	12	0.1000	0.0972 (0.01 min)	0.0998 (0.0060)
$P(\sum_{i=1}^{40} iT_i \geq 458)$	20	0.1006	0.1004 (0.01 min)	0.1004 (0.0060)
$P(\sum_{i=1}^{40} iT_i \geq 375)$	16	0.1009	0.0976 (0.01 min)	0.1006 (0.0060)
$P(\sum_{i=1}^{100} iT_i \geq 2,707)$	50		0.1000 (0.02 min)	0.1058 (0.0061)
$P(\sum_{i=1}^{100} iT_i \geq 2,199)$	40		0.1002 (0.03 min)	0.1047 (0.0062)
$P(\sum_{i=1}^{500} iT_i \geq 64,682)$	250		0.1004 (0.06 min)	0.1016 (0.0059)
$P(\sum_{i=1}^{500} iT_i \geq 52,137)$	200		0.1000 (0.06 min)	0.0992 (0.0059)

enables a p–value estimator that is unaffected by treatment imbalance, choice of score functions and does not rely on large sample approximations as importance sampling.

2.3 Error rates and an example

Using the results in Section 2.2, we can now compare the size and power of randomization tests for conditional, quasi–conditional, and unconditional reference sets. We evaluate power under a location-shift alternative (e.g., Flyer, 1998). The following algorithm can be used to simulate the type I error rate of the upper–tail, α–level, conditional randomization test:

1. Generate a set of data $\boldsymbol{z}$ under the null hypothesis.

2. Generate 2500 sequences from the the conditional reference set where n_1 is the

observed number on treatment 1.

3. For each sequence, evaluate the linear rank statistic using the data in $\boldsymbol{z}$.

4. Order the statistics and choose a quantile S_c corresponding to the nominal α level. This is the estimated critical value of the test.

5. To evaluate the probability of type I error, repeat Steps 1 and 2 $\mathcal{L}$ times , each time obtaining the proportion of statistics exceeding S_c. The average of these proportions is the estimated probability of type I error. Denote this average by $\hat{\alpha}$.

To obtain the power of the test, we use the following algorithm:

1. Generate 2500 sequences from the conditional reference set.

2. For each sequence, apply the location–shift to $\boldsymbol{z}$ to create $\boldsymbol{z}_1$. If the jth assignment is to treatment 1, $j = 1, ..., n$, $z_{1j} = z_j + \delta$, and $z_{1j} = z_j$, otherwise.

3. For each sequence, evaluate the linear rank statistic.

4. The proportion of statistics equal to or exceeding the critical value is an estimate of power.

5. Repeat Steps 1 to 4 $\mathcal{L}$ times to approximate the distribution of power. The average of these proportions is the estimated power of the upper–tail, α level test and δ shift.

We take an example from the Diabetes Complications and Control Trial, using the data found in Table 7.4 of Rosenberger and Lachin (2002, p. 112). We compute the van der Waerden scores for the cholesterol outcome data ($n = 50$) and simulate treatment assignments from a BCD($p = 3/4$). The observed number assigned to treatment

Table 2.6: *Simulated error rates; BCD*(0.75)*;* $n = 50$

Type of test	δ	S_c	$\hat{\alpha}$ (S.D.)	Power (S.D.)
Conditional on $\{N_1(50) = 21\}$	30	5.215	0.0508 (0.0044)	0.8797 (0.0064)
Conditional on $\{N_1(50) = 21 \pm 1\}$	30	5.348	0.0508 (0.0043)	0.8801 (0.0067)
Conditional on $\{N_1(50) = 21 \pm 2\}$	30	5.391	0.0508 (0.0044)	0.8791 (0.0060)
Unconditional	30	5.535	0.0508 (0.0043)	0.8702 (0.0069)

Table 2.7: *Simulated error rates; complete randomization;* $n = 50$

Type of test	δ	S_c	$\hat{\alpha}$ (S.D.)	Power (S.D.)
Conditional on $\{N_1(50) = 21\}$	30	5.388	0.0508 (0.0043)	0.8749 (0.0069)
Conditional on $\{N_1(50) = 21 \pm 1\}$	30	5.397	0.0508 (0.0042)	0.8808 (0.0067)
Conditional on $\{N_1(50) = 21 \pm 2\}$	30	5.402	0.0508 (0.0034)	0.8808 (0.0067)
Unconditional	30	5.402	0.0508 (0.0043)	0.8759 (0.0068)

1 was $n_1 = 21$. Using the naive method for the conditional test, 12,719,423 sequences would have to be computed for each replication to estimate error rates. Using the results of Theorem 2.1, we only need to generate 2500 sequences per replication, making $\mathcal{L} = 1000$ replications quite manageable in estimating error rates. Table 2.6 gives the estimate error rates and standard deviations over 1000 replications for the conditional test, two quasi–conditional tests, and the unconditional test. In Table 2.7, we use complete randomization incorrectly to evaluate error rates. Note that the critical values are different, although for this example error rates are similar.

Chapter 3: Monte Carlo Methods for Conditional Stratified Randomization Tests

We extend the procedures discussed in Chapter 2 to approximate conditional stratified randomization tests and provide some applications.

For the beginning, we address the unconditional stratified test, which leads to the first method of approximating conditional stratified tests. Following stratification on known covariates, the computation of a stratified linear rank test based on the randomization distribution is straightforward by summing the stratum–specific test statistics over independent strata. One Monte Carlo approach for the unconditional test is to simply generate N_u collections of I randomization sequences from the product set $\Omega_{u,1} \times \Omega_{u,2} \times ... \times \Omega_{u,I} = \prod_{i=1}^{I} \Omega_{u,i}$, $i = 1, ..., I$, where I is the total number of strata. These collections are composed of I distinct sequences, each sequence coming from a different $\Omega_{u,i}$. To sample one collection from the product set $\prod_{i=1}^{I} \Omega_{u,i}$ one must sample simultaneously and independently one sequence from each one of the $\Omega_{u,i}$'s. Letting $s_1, s_2, ..., s_I$ be the total number of subjects in each stratum, there are $2^{\sum_{i=1}^{I} s_i}$ such collections in the product set. For a given collection $k = 1, ..., N_u$, the test statistic $S_i = \sum_{j=1}^{s_i} (a_{js_i} - \bar{a}_{s_i}) T_{kij}$ is computed within each stratum i. Following, the stratified statistic $\sum_{i=1}^{I} S_i$ is evaluated. The Monte Carlo estimate of the appropriate p–value is obtained from the approximate randomization distribution of $\sum_{i=1}^{I} S_i$.

Sampling from $\prod_{i=1}^{I} \Omega_{u,i}$ to approximate a conditional stratified test would require a considerably large Monte Carlo sample size. Let N_{1i} be the number of patients

assigned to treatment 1, in the ith stratum, and let m_{1i} be the observed number. With this scheme K^* collections must be generated from $\prod_{i=1}^{I} \Omega_{u,i}$ such that enough of them satisfy $\{N_{11} = m_{11}, ..., N_{1I} = m_{1I}\}$. This is a very restrictive set, and will necessarily require a very large Monte Carlo sample. The random variable K^* is distributed as negative binomial with parameters $\pi = \prod_{i=1}^{I} P(N_{1i} = m_{1i})$ and $r = N_u$. As an example, suppose there are four strata with the following proportions assigned to treatment 1:

$$(6/12,\ 4/10,\ 5/9,\ 4/8).$$

Under the BCD(3/4), the required number of collections is $85,959$ and if the proportions are multiplied by five – $47,029,393,417$. This procedure will become very slow quickly.

Sampling strictly from $\prod_{i=1}^{I} \Omega_{c,i}$ works better for stratified tests, but it can only be applied for the BCD(p) and complete randomization. The idea is the same as before. A sequence is sampled from each $\Omega_{c,i}$ independently. The process is repeated N_c times and $\sum_{i=1}^{I} S_i$ is evaluated each time. The Monte Carlo p–value is obtained from the approximated randomization distribution of $\sum_{i=1}^{I} S_i$. Taking $N_c = 2500$ works as well as in the unstratified case. In Table 3.1, we apply this technique to approximate the upper 0.05 tail of the randomization distribution of the stratified Wilcoxon statistic when having four strata with the above proportions assigned to treatment 1, and for the cases when the proportions are multiplied by $2, 5, 7$ and 10. With four strata and about 100 subjects in each strata, a p–value approximation can be obtained in less than five minutes. The run time would double if there were eight strata instead. Note that these run times are not affected by the value of the probability parameter p in the BCD(p), which is the strength of this sampling scheme. This technique can also be effectively applied to conditional stratified tests under complete randomization.

Table 3.1: *Approximations for the upper* 0.05 *tail of the conditional stratified randomization distribution of the statistic* $\sum_{i=1}^{4}\sum_{j=1}^{s_i} jT_j$ *by sampling from* $\prod_{i=1}^{I}\Omega_{c,i}$*;* $BCD(3/4)$*;* $N_c = 2500$

	$\sum_{i=1}^{4} s_i$	Exact	1 MC run	1000 MC runs; mean (S.D.)
$P(\sum_{i=1}^{4}\sum_{j=1}^{s_i} jT_j \geq 113)$	39	0.0661	0.0672 (0.1 min)	0.0663 (0.0050)
$P(\sum_{i=1}^{4}\sum_{j=1}^{s_i} jT_j \geq 417)$	78		0.0520 (0.3 min)	0.0479 (0.0043)
$P(\sum_{i=1}^{4}\sum_{j=1}^{s_i} jT_j \geq 2450)$	195		0.0504 (1.4 min)	0.0483 (0.0045)
$P(\sum_{i=1}^{4}\sum_{j=1}^{s_i} jT_j \geq 4736)$	273		0.0516 (2.5 min)	0.0554 (0.0046)
$P(\sum_{i=1}^{4}\sum_{j=1}^{s_i} jT_j \geq 9580)$	390		0.0504 (4.8 min)	0.0507 (0.0044)

Approximating conditional stratified randomization tests has not been addressed in the literature so far. We believe the technique that samples directly from the conditional randomization set works reasonably well for this purpose and is the best available solution.

Chapter 4: Monte Carlo Methods for the Sequential Monitoring of Randomization Tests

Sequential monitoring refers to the analysis of the data periodically during the course of a clinical trial, with the purpose of detecting early evidence in support of or against a hypothesis. A desirable feature of such a monitoring plan would be flexible inspections of the data that can occur at arbitrary time points. At the same time, sequentially tested hypotheses must maintain the overall probability of type I error at the prespecified level, since repeated testing is known to inflate it. The Lan and DeMets (1988) error spending approach for sequential monitoring allows for just that. This approach makes use of a type I error spending function, which depends on the amount of "statistical information" available at the time of the interim inspection. In the context of sequential monitoring, the "statistical information" is a measure of how far a trial has progressed. Under a population model, the amount of interim information – the information fraction – is defined as the proportion of Fisher's information observed thus far in the trial. The type I error spending function rations the amount of type I error that may be spent at each look commensurate to the information fraction. The critical value associated with the allowable probability of type I error at a certain interim look is obtained and compared to the observed value of the statistic. The decision whether to continue, or stop, the trial is based on this comparison.

In this chapter we solve the problem of computing unconditional and conditional randomization tests following the BCD(p) and complete randomization when sequential monitoring is implemented in the course of the clinical trial. We provide a Monte

Carlo algorithm for approximating the critical values of such tests. This algorithm uses the idea of sampling from the conditional reference set.

4.1 Sequential monitoring of unconditional randomization tests

Suppose there are $L-1$ interim inspections of the data after $1 \leq r_1 < r_2 < ... < r_{L-1} < r_L = n$ patients responded. Let $0 < t_1 < t_2 < ... < t_{L-1} < t_L = 1$ be the corresponding information fraction at those inspections. Let the linear–rank randomization test statistic computed at each of the inspections be given by $V_{r_l} = \sum_{j=1}^{r_l}(a_{jr_l} - \overline{a}_{r_l})T_j = \boldsymbol{a}'_{r_l}\boldsymbol{T}^{(r_l)},\ l = 1, ..., L$. Using the alpha–spending function approach (Lan and DeMets, 1983), let $\alpha^*(t), t \in [0, 1]$ be a nondecreasing function such that $\alpha^*(0) = 0$ and $\alpha^*(1) = \alpha$, the significance level of the one–sided test. One such function is

$$\alpha^*(t) = \begin{cases} 0 & t = 0, \\ 2 - 2\Phi(z_{\alpha/2}/\sqrt{t}) & 0 < t \leq 1, \end{cases}$$

where Φ is the standard normal distribution function and $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ (Lan and DeMets, 1983; O'Brien and Fleming, 1979). The constants $d_1, ..., d_L$ that satisfy

$$\begin{cases} P(V_{r_1} > d_1) & = \alpha^*(t_1), \\ P(V_{r_1} \leq d_1, V_{r_2} > d_2) & = \alpha^*(t_2) - \alpha^*(t_1), \\ P(V_{r_1} \leq d_1, V_{r_2} \leq d_2, V_{r_3} > d_3) & = \alpha^*(t_3) - \alpha^*(t_2), \\ & \vdots \\ P(V_{r_1} \leq d_1, V_{r_2} \leq d_2, ..., V_{r_{L-1}} \leq d_{L-1}, V_n > d_L) & = \alpha - \alpha^*(t_{L-1}), \end{cases} \tag{4.1}$$

demarcate the proper rejection region of an α–level, upper–tailed, unconditional randomization test with $L-1$ interim inspections. For the sequential monitoring of an α–level, lower–tailed test, all inequalities in (4.1) must be reversed. The problem addressed in this section is that of estimating the quantiles $d_1, ..., d_L$ by Monte Carlo methods.

There are three issues when setting to estimate the quantiles $d_1, ..., d_L$ in (4.1):

1. sampling from discrete multivariate distributions;

2. computing $t_1, t_2, ..., t_L$ in the context of randomization tests;

3. choosing an estimator for quantiles in the tail of discrete distributions.

The best way to deal with the first issue is to avoid sampling from multivariate distributions, by expressing the problem in terms of univariate distributions only. This is possible by noting that the set of conditions in (4.1) is equivalent to the following set of conditions, expressed in terms of univariate conditional distributions

$$\begin{cases} P(V_{r_1} > d_1) & = \alpha^*(t_1), \\ P(V_{r_2} > d_2 | V_{r_1} \leq d_1) & = \frac{\alpha^*(t_2)-\alpha^*(t_1)}{1-\alpha^*(t_1)}, \\ P(V_{r_3} > d_3 | V_{r_1} \leq d_1, V_{r_2} \leq d_2) & = \frac{\alpha^*(t_3)-\alpha^*(t_2)}{1-\alpha^*(t_2)}, \\ & \vdots \\ P(V_n > d_L | V_{r_1} \leq d_1, V_{r_2} \leq d_2, ..., V_{r_{L-1}} \leq d_{L-1}) & = \frac{\alpha-\alpha^*(t_{L-1})}{1-\alpha^*(t_{L-1})}. \end{cases} \tag{4.2}$$

LEMMA 4.1 *The sets of conditions* (4.1) *and* (4.2) *are equivalent.*

PROOF. The first condition is identical for both sets. According to (4.1), the boundary d_1, d_2 at the 2nd inspection must satisfy $P(V_{r_1} \leq d_1, V_{r_2} > d_2) = \alpha^*(t_2) - \alpha^*(t_1)$. Noting that $P(V_{r_1} \leq d_1) = 1 - \alpha^*(t_1)$ by (4.1), the above condition is equivalent to

$$P(V_{r_2} > d_2 | V_{r_1} \leq d_1) = \frac{\alpha^*(t_2) - \alpha^*(t_1)}{P(V_{r_1} \leq d_1)} = \frac{\alpha^*(t_2) - \alpha^*(t_1)}{1 - \alpha^*(t_1)}.$$

Similarly for $l = 3$, the condition $P(V_{r_1} \leq d_1, V_{r_2} \leq d_2, V_{r_3} > d_3) = \alpha^*(t_3) - \alpha^*(t_2)$ is equivalent to

$$P(V_{r_3} > d_3 | V_{r_1} \leq d_1, V_{r_2} \leq d_2) = \frac{\alpha^*(t_3) - \alpha^*(t_2)}{P(V_{r_1} \leq d_1, V_{r_2} \leq d_2)} = \frac{\alpha^*(t_3) - \alpha^*(t_2)}{1 - \alpha^*(t_2)},$$

because by (4.1), $P(V_{r_1} \leq d_1, V_{r_2} > d_2) = [\alpha^*(t_2) - \alpha^*(t_1)]$ and

$$P(V_{r_1} \leq d_1, V_{r_2} \leq d_2) = P(V_{r_1} \leq d_1) - P(V_{r_1} \leq d_1, V_{r_2} > d_2)$$

$$= 1 - \alpha^*(t_1) - [\alpha^*(t_2) - \alpha^*(t_1)] = 1 - \alpha^*(t_2).$$

For $3 \leq l \leq L$, one can show recursively that

$$P(V_{r_1} \leq d_1, V_{r_2} \leq d_2, ..., V_{r_l} \leq d_l) = 1 - \alpha^*(t_l).$$

$$P(V_{r_1} \leq d_1, ..., V_{r_l} \leq d_l) = P(V_{r_1} \leq d_1, ..., V_{r_{l-1}} \leq d_{l-1}) - P\left(\bigcap_{j=1}^{l-1} \{V_{r_j} \leq d_j\}, V_{r_l} > d_l\right)$$

$$= P(V_{r_1} \leq d_1, ..., V_{r_{l-1}} \leq d_{l-1}) - [\alpha^*(t_l) - \alpha^*(t_{l-1})], \text{ by (4.1)}$$

$$= P(V_{r_1} \leq d_1, ..., V_{r_{l-2}} \leq d_{l-2}) - P\left(\bigcap_{j=1}^{l-2} \{V_{r_j} \leq d_j\}, V_{r_{l-1}} > d_{l-1}\right) - [\alpha^*(t_l) - \alpha^*(t_{l-1})]$$

$$= P(V_{r_1} \leq d_1, ..., V_{r_{l-2}} \leq d_{l-2}) - [\alpha^*(t_{l-1}) - \alpha^*(t_{l-2})] - [\alpha^*(t_l) - \alpha^*(t_{l-1})]$$

$$\vdots$$

$$= P(V_{r_1} \leq d_1, V_{r_2} \leq d_2) - [\alpha^*(t_3) - \alpha^*(t_2)] - [\alpha^*(t_4) - \alpha^*(t_3)] - ...$$

$$- [\alpha^*(t_{l-1}) - \alpha^*(t_{l-2})] - [\alpha^*(t_l) - \alpha^*(t_{l-1})]$$

$$= [1 - \alpha^*(t_2)] - [\alpha^*(t_3) - \alpha^*(t_2)] - [\alpha^*(t_4) - \alpha^*(t_3)] - ... - [\alpha^*(t_l) - \alpha^*(t_{l-1})] = 1 - \alpha^*(t_l).$$

Thus for $4 \leq l \leq L$, $P(V_{r_1} \leq d_1, V_{r_2} \leq d_2, ..., V_{r_l} > d_l) = \alpha^*(t_l) - \alpha^*(t_{l-1})$ is equivalent to

$$P(V_{r_l} > d_l | V_{r_1} \leq d_1, V_{r_2} \leq d_2, ..., V_{r_{l-1}} \leq d_{l-1}) = \frac{\alpha^*(t_l) - \alpha^*(t_{l-1})}{P(V_{r_1} \leq d_1, V_{r_2} \leq d_2, ..., V_{r_{l-1}} \leq d_{l-1})}$$

$$= \frac{\alpha^*(t_l) - \alpha^*(t_{l-1})}{1 - \alpha^*(t_{l-1})}. \ \square$$

The new set of conditions (4.2) involves only conditional univariate distributions, which are much easier to sample from than the joint distributions in (4.1). The Monte Carlo algorithm has L stages. Each stage approximates one of the d_l's. Suppose a sample of size N_u is sufficient to estimate a distribution quantile using some quantile estimator. The Monte Carlo algorithm that estimates the boundary $d_1, ..., d_L$ for an α–level, upper–tailed, unconditional randomization test with $L-1$ interim inspections is as follows:

1. At stage 1, generate N_u randomization sequences of r_1 assignments from the appropriate reference set. Evaluate V_{r_1} for each sequence; estimate d_1 using the quantile estimator of choice based on the values of V_{r_1}.

2. At stage 2, generate $N_u/(1-\alpha^*(t_1))$ randomization sequences of r_2 assignments from the appropriate reference set. For each sequence, evaluate V_{r_1} using the first r_1 of r_2 assignments only. Retain those sequences that satisfy $\{V_{r_1} \leq d_1\}$. Evaluate V_{r_2} for each retained sequence. Estimate d_2 using the quantile estimator of choice based on the values of V_{r_2}.

3. At stage 3, generate $N_u/(1-\alpha^*(t_1))\,(1 - [\alpha^*(t_3) - \alpha^*(t_2)]/[1 - \alpha^*(t_2)])$ randomization sequences of r_3 assignments from the appropriate reference set. For each sequence, evaluate V_{r_1} and V_{r_2} using the first r_1 and r_2 assignments, respectively.

Retain those sequences that satisfy $\{V_{r_1} \leq d_1\}$ and $\{V_{r_2} \leq d_2\}$. Evaluate V_{r_3} for each retained sequence. Estimate d_3 using the quantile estimator of choice based on the values of V_{r_3}.

4. At stage $4 \leq l \leq L$, generate $N_u / \prod_{i=1}^{l}(1 - [\alpha^*(t_i) - \alpha^*(t_{i-1})]/[1 - \alpha^*(t_{i-1})])$ randomization sequences of r_l assignments from the unconditional reference set. Note that $\alpha^*(t_0) = 0$ and $\alpha^*(t_L) = \alpha$. For each sequence, evaluate $V_{r_1}, V_{r_2}, ..., V_{r_{l-1}}$ using the first $r_1, r_2, ..., r_{l-1}$ assignments, respectively. Retain those sequences that satisfy $\bigcap_{i=1}^{l-1}\{V_{r_i} \leq d_i\}$. Evaluate V_{r_l} for each retained sequence. Estimate d_l using the quantile estimator of choice based on the values of V_{r_l}.

Requiring that $N_u / \prod_{i=1}^{l}(1 - [\alpha^*(t_i) - \alpha^*(t_{i-1})]/[1 - \alpha^*(t_{i-1})])$ randomization sequences be sampled at stage l simply ensures that at least N_u sequences are used for the estimation of d_l at each stage. Next, we discuss the concept of information in randomization–based sequential tests.

Under a randomization model, the concept of Fisher's information does not exist as under a population model. However, since the Fisher's information approximates the inverse of the asymptotic variance of the test, it seems reasonable to define a randomization–based information fraction as the ratio of the variances (Rosenberger and Lachin, 2002). Let $\boldsymbol{\Sigma}_{r_l} = \text{Var}(\boldsymbol{T}^{(r_l)})$, the covariance matrix of $\boldsymbol{T}^{(r_l)}$. We can write

$$t_l = \frac{\text{Var}\left(\boldsymbol{a}'_{r_l}\boldsymbol{T}^{(r_l)}\right)}{\text{Var}\left(\boldsymbol{a}'_{n}\boldsymbol{T}^{(n)}\right)} = \frac{\boldsymbol{a}'_{r_l}\boldsymbol{\Sigma}_{r_l}\boldsymbol{a}_{r_l}}{\boldsymbol{a}'_{n}\boldsymbol{\Sigma}_{n}\boldsymbol{a}_{n}}. \tag{4.3}$$

For complete randomization and BCD(p), unlike for the GBCD, there exist analytical forms for $\boldsymbol{\Sigma}_n$ (Rosenberger and Lachin, 2002; Markaryan and Rosenberger, 2010), in which case it's possible to evaluate exactly the variance in the numerator of (4.3). In

the denominator, however, $\boldsymbol{a}'_n$ is unknown, since at each interim inspections a portion of the data is unknown. One would have to interpolate sequentially the remaining unknown data points in order to have a value for $\boldsymbol{a}'_n$. In Section 4.3, we examine several methods for data interpolation and their effect on the information.

Quantile estimation for discrete distributions has not been studied as extensively as for continuous distributions. Let $Z_1, ..., Z_{N_u}$ be a random sample from a discrete distribution, F. One may wish to estimate ξ_p, the pth quantile of F $(0 < p < 1)$, by the sample quantile estimator

$$\tilde{\xi}_p = \begin{cases} Z_{(N_u p)} & \text{if } N_u p \in \mathbb{N}, \\ Z_{([N_u p]+1)} & \text{if } N_u p \notin \mathbb{N}, \end{cases} \tag{4.4}$$

where the notation $Z_{(i)}$ is for the ith order statistic and $\mathbb{N}$ is the set of positive integers. The estimator $\tilde{\xi}_p$ has been shown to be consistent for ξ_p only in the case of continuous distributions. Chen and Lazar (2010) formalize a consistent and nonparametric quantile estimator in the case of discrete distributions. Their estimator is a "corrected" version of $\tilde{\xi}_p$ that involves checking a series of conditions. They recommend a sample size of 500 to 10,000 when using their estimator. For simplicity, we use $\tilde{\xi}_p$ with a sample size of 10,000, which in our simulations later on had sufficiently good performance.

4.2 Sequential monitoring of conditional randomization tests

For conditional tests, let $N_1(r_1), N_1(r_2), ..., N_1(r_{L-1})$, and $N_1(r_L) = N_1(n)$ be the sample sizes assigned to treatment 1 after inspections $1, ..., L$ and let $n_{11}, ..., n_{1(L-1)}$,

and $n_{1L} = n_1$ be realizations of these sample sizes. Following Zhang and Rosenberger (2008), the upper–tailed, conditional randomization test with L interim looks involves finding $d_1, ..., d_L$ such that

$$\begin{cases} P(V_{r_1} > d_1 | N_1(r_1) = n_{11}) & = \alpha^*(t_1), \\ P(V_{r_1} \leq d_1, V_{r_2} > d_2 | N_1(r_1) = n_{11}, N_1(r_2) = n_{12}) & = \alpha^*(t_2) - \alpha^*(t_1), \\ P(V_{r_1} \leq d_1, V_{r_2} \leq d_2, V_{r_3} > d_3 | \bigcap_{j=1}^{3} N_1(r_j) = n_{1j}) & = \alpha^*(t_3) - \alpha^*(t_2), \\ & \vdots \\ P(V_{r_1} \leq d_1, ..., V_{r_{L-1}} \leq d_{L-1}, V_n > d_L | \bigcap_{j=1}^{L} N_1(r_j) = n_{1j}) & = \alpha - \alpha^*(t_{L-1}). \end{cases} \tag{4.5}$$

For lower–tailed tests, all inequalities in (4.5) must be reversed. The asymptotic joint normality of these conditional distributions has not been shown, except in the case of $L = 2$ under the GBCD (Zhang and Rosenberger, 2008). Since Efron's BCD(p) does not admit an asymptotically normal linear rank test, these asymptotic techniques cannot be used at all, either conditionally or unconditionally.

The following lemma expresses (4.5) in terms of univariate conditional distributions. The proof makes use of the Markovian property addressed in Lemma 2.3.

LEMMA 4.2 *The set of conditions* (4.5) *is equivalent to*

$$\begin{cases} P\left(V_{r_1} > d_1 | N_1(r_1) = n_{11}\right) & = \alpha^*(t_1), \\ P\left(V_{r_2} > d_2 | V_{r_1} \leq d_1, \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right) & = \frac{\alpha^*(t_2) - \alpha^*(t_1)}{1 - \alpha^*(t_1)}, \\ P\left(V_{r_3} > d_3 | \bigcap_{j=1}^{2} \{V_{r_j} \leq d_j\}, \bigcap_{j=1}^{3} \{N_1(r_j) = n_{1j}\}\right) & = \frac{\alpha^*(t_3) - \alpha^*(t_2)}{1 - \alpha^*(t_2)}, \\ & \vdots \\ P\left(V_n > d_L | \bigcap_{j=1}^{L-1} \{V_{r_j} \leq d_j\}, \bigcap_{j=1}^{L} \{N_1(r_j) = n_{1j}\}\right) & = \frac{\alpha - \alpha^*(t_{L-1})}{1 - \alpha^*(t_{L-1})}. \end{cases} \tag{4.6}$$

PROOF. The first conditions in (4.6) and (4.5) are identical. At the 2nd inspection

in (4.5), d_1, d_2 must satisfy

$$P\left(V_{r_1} \leq d_1, V_{r_2} > d_2 | \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right) = \alpha^*(t_2) - \alpha^*(t_1).$$

The left hand side of this condition can be expressed as

$$P\left(V_{r_1} \leq d_1, V_{r_2} > d_2 | \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right)$$

$$= \frac{P\left(V_{r_1} \leq d_1, V_{r_2} > d_2, \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right)}{P\left(\bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right)}$$

$$= \frac{P\left(V_{r_2} > d_2 | V_{r_1} \leq d_1, \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right) P\left(V_{r_1} \leq d_1 | \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right)}{P\left(\bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right)}$$

$$\times P\left(\bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right)$$

$$= P\left(V_{r_1} \leq d_1 | \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right) P\left(V_{r_2} > d_2 | V_{r_1} \leq d_1, \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right)$$

$$= (1 - \alpha^*(t_1)) P\left(V_{r_2} > d_2 | V_{r_1} \leq d_1, \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right), \text{ where}$$

$$P\left(V_{r_1} \leq d_1 | \bigcap_{j=1}^{2} \{N_1(r_j) = n_{1j}\}\right) = \frac{P\left(N_1(r_2) = n_{12} | V_{r_1} \leq d_1, N_1(r_1) = n_{11}\right)}{P(N_1(r_1) = n_{11}) P(N_1(r_2) = n_{12} | N_1(r_1) = n_{11})}$$

$$\times P\left(V_{r_1} \leq d_1, N_1(r_1) = n_{11}\right)$$

$$= P\left(V_{r_1} \leq d_1 | N_1(r_1) = n_{11}\right), \text{ by Lemma 2.3}$$

$$= 1 - \alpha^*(t_1), \text{ by (4.5)}$$

Combined with the right hand side of condition two in (4.5), the boundary d_1, d_2 must satisfy

$$P\left(V_{r_2} > d_2 | V_{r_1} \leq d_1, \bigcap_{j=1}^{2}\{N_1(r_j) = n_{1j}\}\right) = \frac{\alpha^*(t_2) - \alpha^*(t_1)}{1 - \alpha^*(t_1)}.$$

In general, by Lemma 2.3, for $2 \leq l \leq L-1$,

$$P\left(V_{r_1} \leq d_1, ..., V_{r_l} \leq d_l | \bigcap_{j=1}^{l+1}\{N_1(r_j) = n_{1j}\}\right) = P\left(V_{r_1} \leq d_1, ..., V_{r_l} \leq d_l | \bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right).$$

Then, as in the unconditional case one can show recursively that

$$P\left(V_{r_1} \leq d_1, ..., V_{r_l} \leq d_l | \bigcap_{j=1}^{l+1} N_1(r_j) = n_{1j}\right) = 1 - \alpha^*(t_l), \quad 2 \leq l \leq L-1,$$

using (4.5). Finally, for $3 \leq l \leq L$, the left hand side of

$$P\left(\bigcap_{j=1}^{l-1}\{V_{r_j} \leq d_j\}, V_{r_l} > d_l | \bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right) = \alpha^*(t_l) - \alpha^*(t_{l-1})$$

in (4.5), is equivalent to

$$\begin{aligned}
&\frac{P\left(\bigcap_{j=1}^{l-1}\{V_{r_j} \leq d_j\}, V_{r_l} > d_l, \bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right)}{P\left(\bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right)} \\
&= P\left(V_{r_l} > d_l | \bigcap_{j=1}^{l-1}\{V_{r_j} \leq d_j\}, \bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right) \\
&\times \frac{P\left(\bigcap_{j=1}^{l-1}\{V_{r_j} \leq d_j\}, \bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right)}{P\left(\bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right)} \\
&= P\left(V_{r_l} > d_l | \bigcap_{j=1}^{l-1}\{V_{r_j} \leq d_j\}, \bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right)\left(1 - \alpha^*(t_{l-1})\right),
\end{aligned}$$

after applying Bayes Theorem and the earlier recursive result at the third line. It follows that

$$P\left(V_{r_l} > d_l | \bigcap_{j=1}^{l-1}\{V_{r_j} \leq d_j\}, \bigcap_{j=1}^{l}\{N_1(r_j) = n_{1j}\}\right) = [\alpha^*(t_l) - \alpha^*(t_{l-1})]/[1 - \alpha^*(t_{l-1})]. \ \square$$

For the purpose of estimating $d_1, d_2, ..., d_L$, (4.6) allows using the same Monte Carlo algorithm as the one described in Section 4.1, provided that at each stage the allocation sequences are sampled from the appropriate conditional reference set. At each stage l, the conditional reference set is the collection of all sequences satisfying $\bigcap_{i=1}^{l}\{N_1(r_i) = n_{1i}\}$. The following theorem can be used to sample sequences from such sets.

THEOREM 4.1. *Let $1 \leq l \leq L$, $r_0, r_1, r_2, ..., r_l$ and $n_{10}, n_{11}, ..., n_{1l}$ be defined as before, with $r_0 = 0$ and $n_{10} = 0$. Let $k = 1, ..., l$. For $r_{k-1} \leq j < r_k$, $n_{1(k-1)} \leq m_j \leq j$ and $\phi_{j+1}(m_j) = P(T_{j+1} = 1 | \sum_{i=1}^{j} T_i = m_j)$, the rule*

$$\psi_{j+1} = \phi_{j+1}(m_j)\frac{P(N_1(r_k) = n_{1k}|N_1(j+1) = m_j + 1)}{P(N_1(r_k) = n_{1k}|N_1(j) = m_j)} \tag{4.7}$$

can be used to sample a sequence that satisfies $\bigcap_{i=1}^{l}\{N_1(r_i) = n_{1i}\}$.

PROOF. The reference set of sequences satisfying $\bigcap_{i=1}^{l}\{N_1(r_i) = n_{1i}\}$ is a subset of the reference set satisfying $N_1(r_l) = n_{1l}$. Sampling from the latter set was addressed in Chapter 2, by conditioning on $N_1(r_l) = n_{1l}$ at each step in the restricted randomization rule. Likewise, conditioning on $\bigcap_{i=1}^{l}\{N_1(r_i) = n_{1i}\}$ in the restricted randomization rule allows sampling from the set satisfying $\bigcap_{i=1}^{l}\{N_1(r_i) = n_{1i}\}$. Let $l = 1, ..., L$ and $k = 1, ..., l$. When $k = 1$, $0 \leq j < r_1$ and so

$$\psi_{j+1} = P\left(T_{j+1} = 1 | \sum_{i=1}^{j} T_i = m_j, \bigcap_{i=1}^{l} \{N_1(r_i) = n_{1i}\}\right)$$

$$= \frac{P(\sum_{i=1}^{j} T_i = m_j) P(T_{j+1} = 1 | \sum_{i=1}^{j} T_i = m_j) P(N_1(r_1) = n_{11} | N_1(j+1) = m_j + 1)}{P(\sum_{i=1}^{j} T_i = m_j) P(N_1(r_1) = n_{11} | N_1(j) = m_j)}$$

$$\times \frac{\prod_{i=2}^{l} P\left(N_1(r_i) = n_{1i} | N_1(j+1) = m_j + 1, \bigcap_{q=1}^{i-1} \{N_1(r_q) = n_{1q}\}\right)}{\prod_{i=2}^{l} P\left(N_1(r_i) = n_{1i} | N_1(j) = m_j, \bigcap_{q=1}^{i-1} \{N_1(r_q) = n_{1q}\}\right)}$$

$$= \phi_{j+1}(m_j) \frac{P(N_1(r_1) = n_{11} | N_1(j+1) = m_j + 1)}{P(N_1(r_1) = n_{11} | N_1(j) = m_j)},$$

after an application of Lemma 2.3 in the second to last fraction. Note that this result is the same as (2.6) with $n = r_1$ and $n_1 = n_{11}$. For $k = 2, ..., l$ and $r_{k-1} \leq j < r_k$ it follows that

$$\psi_{j+1} = P\left(T_{j+1} = 1 | \sum_{i=1}^{j} T_i = m_j, \bigcap_{i=1}^{l} \{N_1(r_i) = n_{1i}\}\right)$$

$$= \frac{P\left(T_{j+1} = 1, \sum_{i=1}^{j} T_i = m_j, \bigcap_{i=1}^{l} \{N_1(r_i) = n_{1i}\}\right)}{P\left(\sum_{i=1}^{j} T_i = m_j, \bigcap_{i=1}^{l} \{N_1(r_i) = n_{1i}\}\right)}$$

$$= \frac{P(N_1(r_1) = n_{11}) \prod_{i=2}^{k-1} P\left(N_i(r_i) = n_{1i} | \bigcap_{q=1}^{i-1} \{N_1(r_q) = n_{1q}\}\right)}{P(N_1(r_1) = n_{11}) \prod_{i=2}^{k-1} P\left(N_i(r_i) = n_{1i} | \bigcap_{q=1}^{i-1} \{N_1(r_q) = n_{1q}\}\right)}$$

$$\times \frac{P\left(\sum_{i=1}^{j} T_i = m_j | \bigcap_{i=1}^{k-1} \{N_1(r_i) = n_{1i}\}\right)}{P\left(\sum_{i=1}^{j} T_i = m_j | \bigcap_{i=1}^{k-1} \{N_1(r_i) = n_{1i}\}\right)}$$

$$\times \frac{P\left(T_{j+1} = 1 | N_1(j) = m_j, \bigcap_{i=1}^{k-1} \{N_1(r_i) = n_{1i}\}\right)}{P\left(N_1(r_k) = n_{1k} | N_1(j) = m_j, \bigcap_{i=1}^{k-1} \{N_1(r_i) = n_{1i}\}\right)}$$

$$\times \frac{P\left(N_1(r_k) = n_{1k} | N_1(j+1) = m_j + 1, \bigcap_{i=1}^{k-1} \{N_1(r_i) = n_{1i}\}\right)}{\prod_{i=k+1}^{l} P\left(N_1(r_i) = n_{1i} | N_1(j) = m_j, \bigcap_{q=1}^{i-1} \{N_1(r_q) = n_{1q}\}\right)}$$

$$\times \prod_{i=k+1}^{l} P\left(N_1(r_i) = n_{1i} | N_1(j+1) = m_j + 1, \bigcap_{q=1}^{i-1} \{N_1(r_q) = n_{1q}\}\right)$$

$$= P\left(T_{j+1}=1|N_1(j)=m_j,\bigcap_{i=1}^{k-1}\{N_1(r_i)=n_{1i}\}\right)$$

$$\times\frac{P\left(N_1(r_k)=n_{1k}|N_1(j+1)=m_j+1,\bigcap_{i=1}^{k-1}\{N_1(r_i)=n_{1i}\}\right)}{P\left(N_1(r_k)=n_{1k}|N_1(j)=m_j,\bigcap_{i=1}^{k-1}\{N_1(r_i)=n_{1i}\}\right)}$$

$$\times\frac{\prod_{i=k+1}^{l}P\left(N_1(r_i)=n_{1i}|N_1(j+1)=m_j+1,\bigcap_{q=1}^{i-1}\{N_1(r_q)=n_{1q}\}\right)}{\prod_{i=k+1}^{l}P\left(N_1(r_i)=n_{1i}|N_1(j)=m_j,\bigcap_{q=1}^{i-1}\{N_1(r_q)=n_{1q}\}\right)}$$

$$=\phi_{j+1}(m_j)\frac{P(N_1(r_k)=n_{1k}|N_1(j+1)=m_j+1)}{P(N_1(r_k)=n_{1k}|N_1(j)=m_j)}.$$

The end result follows from an application of Lemma 2.3 (given that $r_{k-1}\leq j<r_k$) at the numerators and denominators of the two fractions preceding the last equality. (Note that when $k=2$, $\prod_{i=2}^{k-1}P\left(N_1(r_i)=n_{1i}|\bigcap_{q=1}^{i-1}\{N_1(r_q)=n_{1q}\}\right)$ would simply not appear in the proof, but everything else would remain the same.) □

For complete randomization, $\psi_{j+1}=(n_{1k}-m_j)/(r_k-j)$, $l=1,...,L$, $k=1,...,l$, $r_{k-1}\leq j<r_k$ and $n_{1(k-1)}\leq m_j\leq j$. For the BCD(p) the numerator and the denominator of ψ_{j+1} must be evaluated according to Theorem 2.1. To obtain a sequence from the reference set satisfying $\bigcap_{i=1}^{l}\{N_1(r_i)=n_{1i}\}$, the sampling must be done in $k=1,...,l$ steps. The following sub–algorithm must be applied at each stage l in the algorithm of Section 4.1 in order to sample a sequence from the appropriate reference set.

1. At stage $k=1$, apply ψ_{j+1} with $r_0\leq j<r_1$ to sample the first r_1 assignments.

2. At stage $k=2$, apply ψ_{j+1} with $r_1\leq j<r_2$ to sample the next r_2-r_1 assignments.

3. At stage $3\leq k\leq l$, apply ψ_{j+1} with $r_{k-1}\leq j<r_k$ to sample the next r_k-r_{k-1} assignments.

We point out that unlike in the case of unconditional test with sequential monitoring, this sampling scheme cannot be applied to conditional tests under Smith's (1984) design, with or without sequential monitoring, since the exact theory for this type of randomization procedure has not been developed yet. We now address the evaluation of the information fractions.

Similarly to the unconditional case, we define the information fractions as a ratio of test variances at the time of the interim analysis. Let $\boldsymbol{\Sigma}_{|r_l} = Var(\boldsymbol{T}^{(r_l)}|N_1(r_1) = n_{11}, ..., N_1(r_l) = n_{1l})$. Then

$$t_l = \frac{\boldsymbol{a}'_{r_l}\boldsymbol{\Sigma}_{|r_l}\boldsymbol{a}_{r_l}}{\boldsymbol{a}'_n\boldsymbol{\Sigma}_{|n}\boldsymbol{a}_n}.$$

Below we provide the exact expression for $\boldsymbol{\Sigma}_{|r_l}$ under complete randomization and the BCD(p). In fact, this expression applies to all restricted randomization procedures defined by (2.4). We assume the following conventions throughout the next developments. For any integers $a \geq b \geq 0$ and $c \geq d \geq 0$,

1. $P(N_1(a) = b|N_1(0) = 0) = P(N_1(a) = b)$.

2. If $a = c$ and $b = d$, $P(N_1(a) = b|N_1(c) = d) = 1$.

3. If $d > b$ or $a - c < b - d$, $P(N_1(a) = b|N_1(c) = d) = 0$.

THEOREM 4.2. *Let $1 \leq l \leq L$, $r_0, r_1, r_2, ..., r_l$ and $n_{10}, n_{11}, ..., n_{1l}$ be defined as before, with $r_0 = 0$ and $n_{10} = 0$. Let $k = 1, ..., l$ and $\phi_i(a) = P(T_i = 1|N_1(i-1) = a)$. For $k = 1$, $0 < i \leq r_1$,*

$$E\left(T_i|\bigcap_{q=1}^{l}\{N_1(r_q) = n_{1q}\}\right) = \frac{\sum_{a=0}^{i-1}\phi_i(a)P(N_1(i-1) = a)P(N_1(r_1) = n_{11}|N_1(i) = a+1)}{P(N_1(r_1) = n_{11})},$$

where $E\left(T_1|\bigcap_{q=1}^{l}\{N_1(r_q) = n_{1q}\}\right) = P(T_1 = 1)P(N_1(r_1) = n_{11}|N_1(1) = 1)/P(N_1(r_1) = n_{11})$.

For $2 \leq k \leq l$, $r_{k-1} < i \leq r_k$, $E\left(T_i | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right)$

$$= \frac{\sum_{a=n_{1(k-1)}}^{i-1} \phi_i(a) P(N_1(i-1) = a | N_1(r_{k-1}) = n_{1(k-1)}) P(N_1(r_k) = n_{1k} | N_1(i) = a+1)}{P(N_1(r_k) = n_{1k} | N_1(r_{k-1}) = n_{1(k-1)})}.$$

For the BCD(p), the conditional probabilities in $E\left(T_i | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right)$ *are given by Theorem 2.1.*

PROOF. The result follows from an application of Bayes Theorem and Lemma 2.3 to $P\left(T_i = 1 | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right)$. □

COROLLARY 4.1. *Let* $1 \leq l \leq L$, $r_0, r_1, r_2, ..., r_l$ *and* $n_{10}, n_{11}, ..., n_{1l}$ *be defined as before, with* $r_0 = 0$ *and* $n_{10} = 0$. *Let* $k = 1, ..., l$ *and* $r_{k-1} < i \leq r_k$. *In case of complete randomization*

$$E\left(T_i | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right) = \frac{n_{1k} - n_{1(k-1)}}{r_k - r_{k-1}}.$$

PROOF. Using Theorem 4.2, for $k = 1$ and $i = 1$,

$$E\left(T_1 | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right) = \frac{1}{2} P(N_1(r_1) = n_{11} | N_1(1) = 1) / P(N_1(r_1) = n_{11})$$

$$= \frac{1}{2} \frac{\binom{r_1 - 1}{n_{11} - 1} (1/2)^{r_1 - 1}}{\binom{r_1}{n_{11}} (1/2)^{r_1}} = \frac{n_{11}}{r_1}.$$

For $k = 1$ and $1 < i \leq r_1$

$$E\left(T_i | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right) = \frac{\sum_{q=0}^{i-1} \phi_i(q) P(N_1(i-1) = q) P(N_1(r_1) = n_{11} | N_1(i) = q+1)}{P(N_1(r_1) = n_{11})}$$

$$= \frac{P\left(T_i = 1, N_1(r_1) = n_{11}\right)}{P(N_1(r_1) = n_{11})} = \frac{n_{11}}{r_1}.$$

For $k = 2, ..., l$ and $r_{k-1} < i \leq r_k$

$$E\left(T_i | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right)$$

$$= \frac{\sum_{q=n_{1(k-1)}}^{i-1} \phi_i(q) P(N_1(i-1) = q | N_1(r_{k-1}) = n_{1(k-1)}) P(N_1(r_k) = n_{1k} | N_1(i) = q+1)}{P(N_1(r_k) = n_{1k} | N_1(r_{k-1}) = n_{1(k-1)})}$$

$$= \frac{P(T_i = 1, N_1(r_{k-1}) = n_{1(k-1)}, N_1(r_k) = n_{1k})}{P(N_1(r_{k-1}) = n_{1(k-1)}) P(N_1(r_k) = n_{1k} | N_1(r_{k-1}) = n_{1(k-1)})}$$

$$= \frac{P(T_i = 1, N_1(r_k) = n_{1k} | N_1(r_{k-1}) = n_{1(k-1)})}{P(N_1(r_k) = n_{1k} | N_1(r_{k-1}) = n_{1(k-1)})}$$

$$= \frac{1/2 \binom{r_k - r_{k-1} - 1}{n_{1k} - n_{1(k-1)} - 1} (1/2)^{r_k - r_{k-1} - 1}}{\binom{r_k - r_{k-1}}{n_{1k} - n_{1(k-1)}} (1/2)^{r_k - r_{k-1}}}$$

$$= \frac{n_{1k} - n_{1(k-1)}}{r_k - r_{k-1}}. \ \square$$

THEOREM 4.3. *Let $1 \leq l \leq L$, $r_0, r_1, r_2, ..., r_l$ and $n_{10}, n_{11}, ..., n_{1l}$ be defined as before, with $r_0 = n_{10} = 0$. Let $k = 1, ..., l$, $\phi_i(a) = P(T_i = 1 | N_1(i-1) = a)$ and $\lambda(a) = \phi_i(a) P(N_1(i-1) = a | N_1(r_{k-1}) = n_{1(k-1)})$. For $1 \leq k \leq l$ and $r_{k-1} < i < j \leq r_k$,*

$$E\left(T_i T_j | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right)$$

$$= \frac{\sum_{a=n_{1(k-1)}}^{i-1} \lambda(a) \sum_{b=n_{1(k-1)+1}}^{j-1} \phi_j(b) P(N_1(j-1) = b | N_1(i) = a+1) P(N_1(r_k) = n_{1k} | N_1(j) = b+1)}{P(N_1(r_k) = n_{1k} | N_1(r_{k-1}) = n_{1(k-1)})}.$$

For all other i, j,

$$E\left(T_i T_j | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right) = E\left(T_i | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right) E\left(T_j | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\}\right).$$

For the BCD(p), the conditional probabilities in $E\left(T_iT_j|\bigcap_{q=1}^{l}\{N_1(r_q)=n_{1q}\}\right)$ *are given by Theorem 2.1.*

PROOF. The result follows from an application of Bayes Theorem and Lemma 2.3 to $P\left(T_i=1,T_j=1|\bigcap_{q=1}^{l}\{N_1(r_q)=n_{1q}\}\right)$. □

COROLLARY 4.2. *Let* $r_0,r_1,r_2,...,r_l$ *be defined as before, with* $r_0=0$, $1\leq l\leq L$, $k=1,...,l$ *and* $1\leq i\leq j\leq r_l$. *Let* $\vartheta_{i|r_l}=E\left(T_i|\bigcap_{q=1}^{l}\{N_1(r_q)=n_{1q}\}\right)$. *The* (i,j)*th entry of* $\mathbf{\Sigma}_{|r_l}$ *under the BCD(p) is*

$$\sigma_{ij}=\begin{cases}E\left(T_iT_j|\bigcap_{q=1}^{l}\{N_1(r_q)=n_{1q}\}\right)-\vartheta_{i|r_l}\vartheta_{j|r_l}, & \text{if}\quad i<j\quad\text{and}\quad r_{k-1}<i<j\leq r_k,\\ \vartheta_{i|r_l}(1-\vartheta_{i|r_l}), & \text{if}\quad i=j,\\ 0, & \text{otherwise,}\end{cases}$$

where the expression for $E\left(T_iT_j|\bigcap_{q=1}^{l}\{N_1(r_q)=n_{1q}\}\right)$ *is given by Theorem 4.3 and the expression for* $\vartheta_{i|r_l}$ *is given by Theorem 4.2.*

COROLLARY 4.3. *Let* $r_0,r_1,r_2,...,r_l$ *and* $n_{10},n_{11},...,n_{1l}$ *be defined as before, with* $r_0=n_{10}=0$. *Let* $1\leq l\leq L$, $k=1,...,l$ *and* $1\leq i\leq j\leq r_l$. *Under complete randomization,* $\mathbf{\Sigma}_{|r_l}$ *is a block diagonal matrix whose* (i,j)*th entry is*

$$\sigma_{ij}=\begin{cases}-\frac{(n_{1k}-n_{1(k-1)})\{r_k-r_{k-1}-(n_{1k}-n_{1(k-1)})\}}{(r_k-r_{k-1})^2(r_k-r_{k-1}-1)}, & \text{if}\quad i<j\quad\text{and}\quad r_{k-1}<i<j\leq r_k,\\ \frac{(n_{1k}-n_{1(k-1)})\{r_k-r_{k-1}-(n_{1k}-n_{1(k-1)})\}}{(r_k-r_{k-1})^2}, & \text{if}\quad i=j,\\ 0, & \text{otherwise.}\end{cases}$$

We note that if $r_k = 2n_{1k}$, for all $1 \leq k \leq l$, the (i,j)th entry of $\boldsymbol{\Sigma}_{|r_l}$ in Corollary 4.3 becomes

$$\sigma_{ij} = \begin{cases} -\frac{1}{4(r_k - r_{k-1} - 1)}, & \text{if} \quad i < j \quad \text{and} \quad r_{k-1} < i < j \leq r_k, \\ \frac{1}{4}, & \text{if} \quad i = j, \\ 0, & \text{otherwise.} \end{cases}$$

In this case, each block k of $\boldsymbol{\Sigma}_{|r_l}$ corresponds to the covariance matrix under the random allocation rule with a sample size of $r_k - r_{k-1}$.

4.3 Sequential monitoring of two–sided randomization tests

If the distribution of the test statistic is symmetric, the generalization of a sequential monitoring plan to a two–sided test is immediate. One simply replaces α with $\alpha/2$, obtains the boundary as in a one–sided test and applies it symmetrically (Lan and DeMets, 1983). The randomization distribution of V_{r_l} may, however, be asymmetric, and in this case the above approach will not lead to quantiles that correspond to the required probability of type I error. Instead, one can take the approach of finding quantiles that are equidistant from the distribution mean. For conditional randomization tests, let

$$\mu_{|r_l} = E\left(V_{r_l} | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\right) = \boldsymbol{a}'_{r_l} E\left(\boldsymbol{T}^{(r_l)} | \bigcap_{q=1}^{l} \{N_1(r_q) = n_{1q}\right),$$

where the ith element of $E\left(\boldsymbol{T}^{(r_l)}|\bigcap_{q=1}^{l}\{N_1(r_q)=n_{1q}\}\right)$ is given by Theorem 4.2 for the BCD(p) and Corollary 4.1 for complete randomization. The constants $d_1, d_2, ..., d_L$ that satisfy

$$\begin{cases} P\left(|V_{r_1}-\mu_{|r_1}|>d_1\middle|\, N_1(r_1)=n_{11}\right) & =\alpha^*(t_1), \\ P\left(|V_{r_2}-\mu_{|r_2}|>d_2\middle|\, |V_{r_1}-\mu_{|r_1}|\le d_1, \bigcap_{j=1}^{2}\{N_1(r_j)=n_{1j}\}\right) & =\frac{\alpha^*(t_2)-\alpha^*(t_1)}{1-\alpha^*(t_1)}, \\ P\left(|V_{r_3}-\mu_{|r_l}|>d_3|\bigcap_{j=1}^{2}\{|V_{r_j}-\mu_{|r_j}|\le d_j\}, \bigcap_{j=1}^{3}\{N_1(r_j)=n_{1j}\}\right) & =\frac{\alpha^*(t_3)-\alpha^*(t_2)}{1-\alpha^*(t_2)}, \\ & \vdots \\ P\left(|V_n-\mu_{|r_L}|>d_L\middle|\, \bigcap_{j=1}^{L-1}\{|V_{r_j}-\mu_{|r_j}|\le d_j\}, \bigcap_{j=1}^{L}\{N_1(r_j)=n_{1j}\}\right) & =\frac{\alpha-\alpha^*(t_{L-1})}{1-\alpha^*(t_{L-1})}, \end{cases}$$

demarcate the rejection region for an α–level, two–sided, conditional randomization test with $L-1$ interim inspections. For unconditional tests, the expectation of the centered sequential statistic is zero under both complete randomization and the BCD(p), since the expectation of the ith assignment is a constant, $E(T_i)=1/2$. Hence, one can simply replace V_{r_l} with $|V_{r_l}|$ in (4.2) for the sequential monitoring of two–sided, unconditional randomization tests with centered scores.

4.4 An example: approximating the information and error rates

We apply the results of Section 4.2 to approximate the information fraction and error rates in an example. The expression for the information fraction was given by $t_l=\boldsymbol{a}'_{r_l}\boldsymbol{\Sigma}_{|r_l}\boldsymbol{a}_{r_l}/\boldsymbol{a}'_n\boldsymbol{\Sigma}_{|n}\boldsymbol{a}_n$. We pointed out in Section 4.1 that $\boldsymbol{a}'_n$ in the denominator of t_l is unknown, since at each inspection a portion of the data is unknown. We examine three methods of interpolating the unknown data. Initially, we generate data for two different groups by sampling $X_i\sim N(1.35, 0.9)$, $i=1,...,126$, and

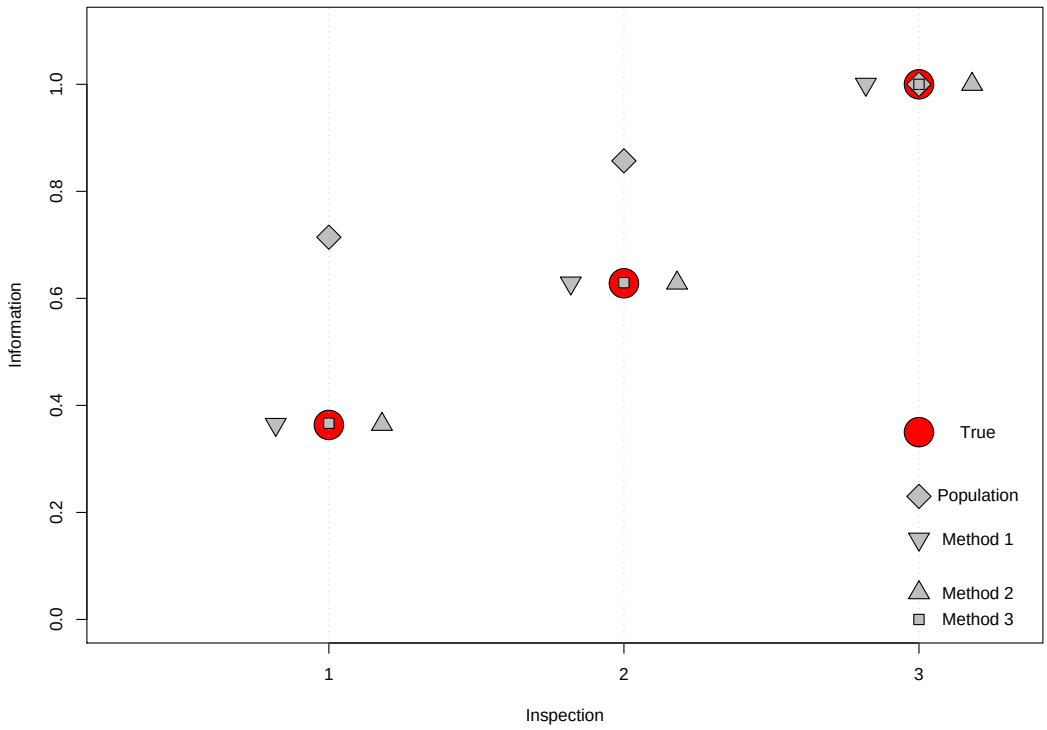


Figure 4.1a: *Information estimation with three interim inspections: complete randomization and no trend in the data,* $L = 3$*;* $r_1 = 250$ *and* $n_{11} = 126$*,* $r_2 = 300$ *and* $n_{12} = 148$*,* $n = 350$ *and* $n_1 = 174$*.*

$Y_i \sim N(1, 0.9),\ i = 1, ..., 124$. It is assumed that the randomization sequence and the total number of subjects at the end of the trial are known in advance, $n = 350$. The data is to be inspected after 250 observations, 300 and 350. At the first inspection the remaining unknown 100 observations are obtained by

1. bootstrapping from the known data by group (Method 1);

2. sampling from a different normal distribution for each group and estimating μ and σ from past data, by group (Method 2);

3. sampling from the same normal distribution and estimating the normal parameters from past data (Method 3).

We compute t_1 under these scenarios and report the results in Figure 4.1a and Table 4.1 for complete randomization and Figure 4.2a and Table 4.1 for the BCD(3/4). At the second inspection, to simulate the progression of the trial, we generate more data by sampling 22 observations from $N(1.35, 0.9)$, and 28 from $N(1, 0.9)$. The remaining unknown 50 observations are obtained with the above methods and t_2 is computed under each one of them. At the third inspection the entire data is available and so it was possible to compute the the true values of $t_l,\ l = 1, 2, 3$ in this example. We also record the population version of the information for comparison, which does not require knowledge of $\boldsymbol{a}'_n$. In the population case $t_l = (n_{1l}(r_l - n_{1l})/r_l)/(n_1(n - n_1)/n)$ (Rosenberger and Lachin, 2002). For both complete randomization and the BCD(p), bootstrapping the unknown data (Method 1) resulted in very accurate results for the randomization based information.

We repeat this procedure by sampling observations from data with a nonlinear trend – $X_i \sim N(0.1\sqrt{i} + 1.35, 0.9)$ and $Y_i \sim N(0.1\sqrt{i} + 1, 0.9)$ – and data with a linear trend – $X_i \sim N(0.1i + 1.35, 0.9)$ and $Y_i \sim N(0.1i + 1, 0.9)$. The results for the data with a nonlinear trend are reported in Figure 4.1b (complete randomization) and

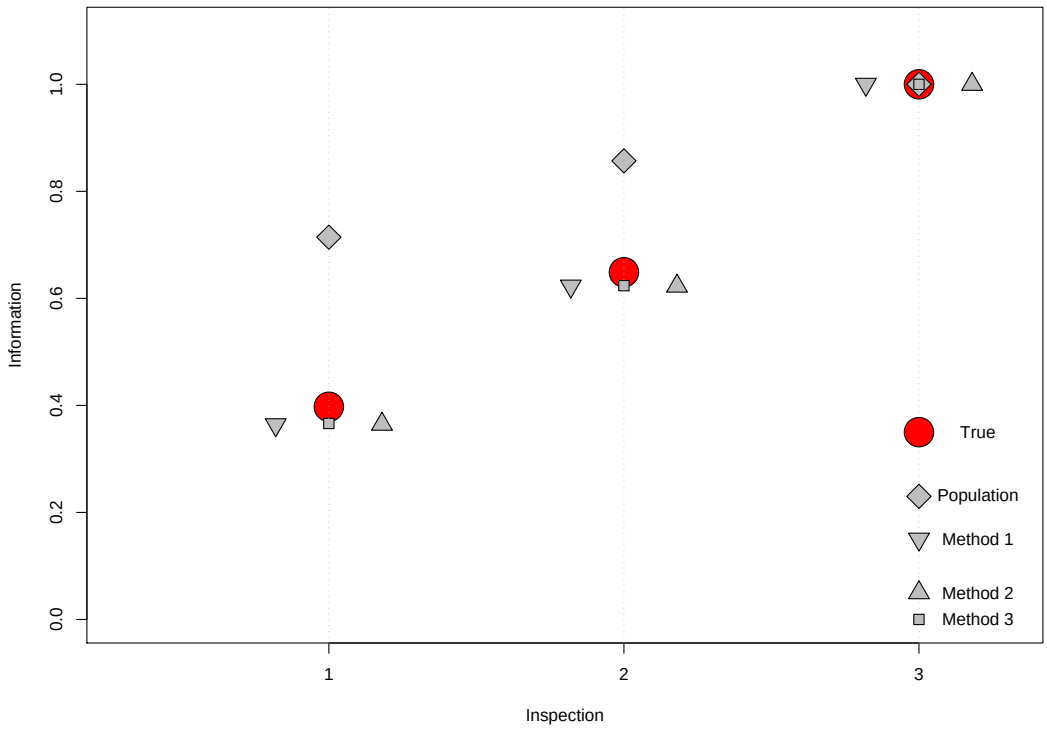


Figure 4.1b: *Information estimation with three interim inspections: complete randomization and nonlinear trend in the data,* $L = 3$*;* $r_1 = 250$ *and* $n_{11} = 126$*,* $r_2 = 300$ *and* $n_{12} = 148$*,* $n = 350$ *and* $n_1 = 174$*.*

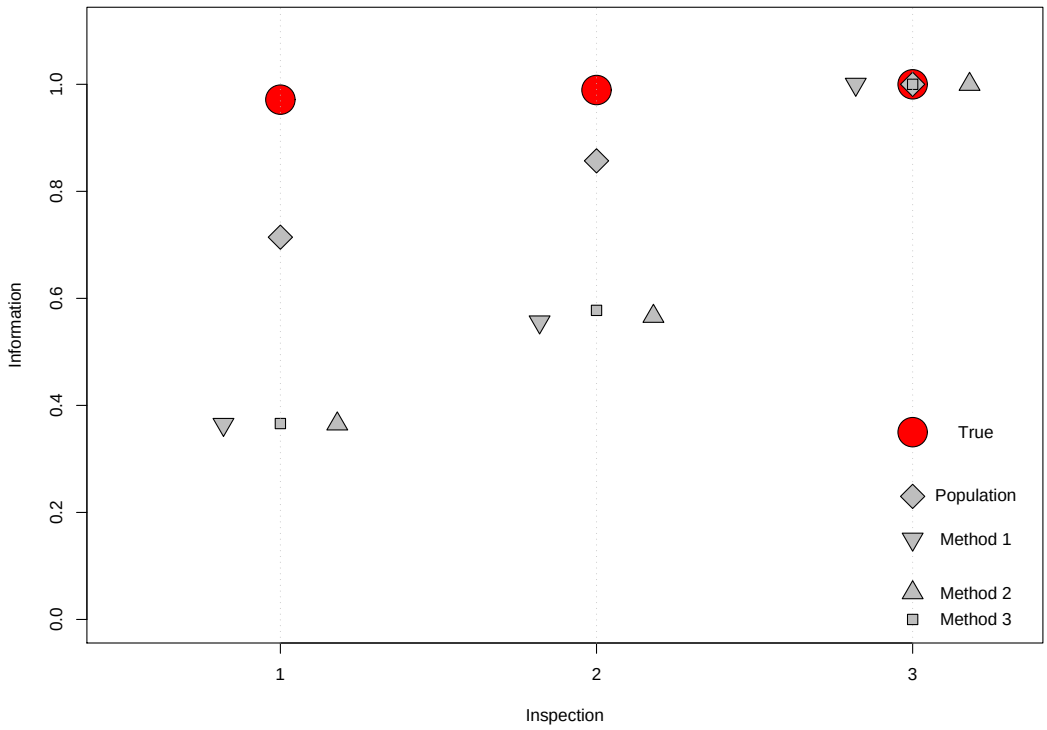


Figure 4.1c: *Information estimation with three interim inspections: complete randomization and linear trend in the data,* $L = 3$*;* $r_1 = 250$ *and* $n_{11} = 126$*,* $r_2 = 300$ *and* $n_{12} = 148$*,* $n = 350$ *and* $n_1 = 174$*.*

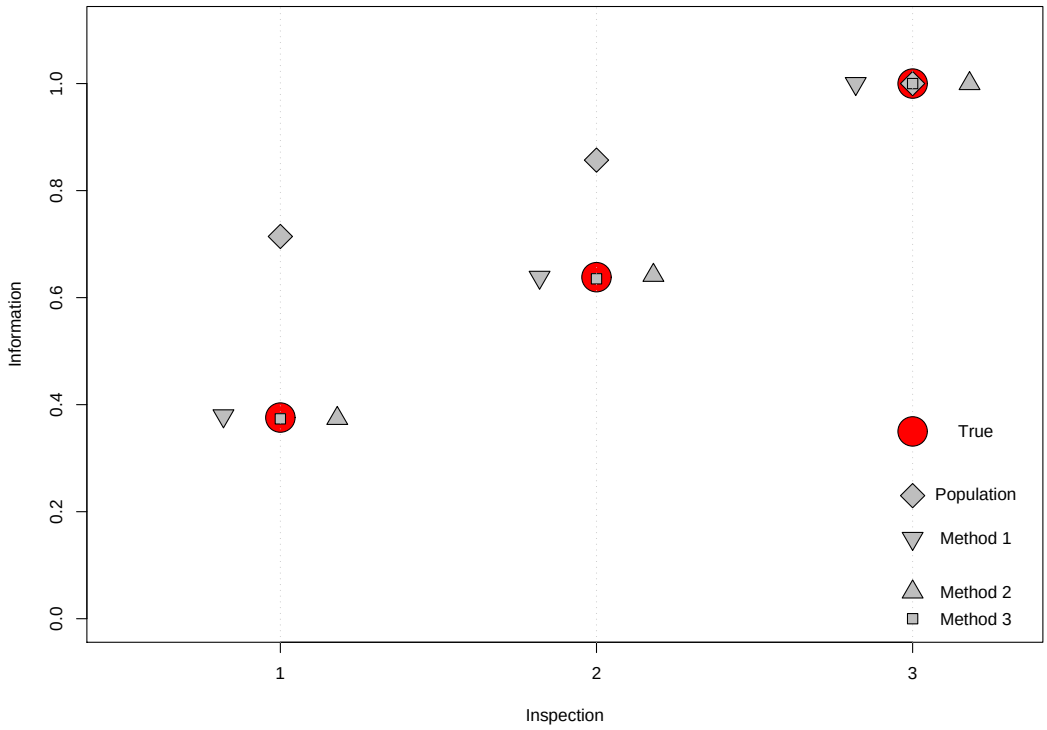


Figure 4.2a: *Information estimation with three interim inspections: BCD(3/4) and no trend in the data,* $L = 3$*;* $r_1 = 250$ *and* $n_{11} = 126$*,* $r_2 = 300$ *and* $n_{12} = 148$*,* $n = 350$ *and* $n_1 = 174$*.*

Figure 4.2b (BCD(3/4)). The results for the data with a linear trend are reported in Figure 4.1c (complete randomization) and Figure 4.2c (BCD(3/4)). In the case of the nonlinear trend, bootstrapping still performs well. In the case of the linear trend, bootstrapping fails and more so for the BCD(p), Figure 4.2c. Computing the information according to the population model gave closer values to the truth for t_l in the latter case.

We also evaluate power under a location–shift alternative (e.g., Flyer, 1998) for a sequential conditional randomization test. The following algorithm can be used to simulate the type I error rate of an upper–tail, α–level, sequential conditional randomization test with L interim looks:

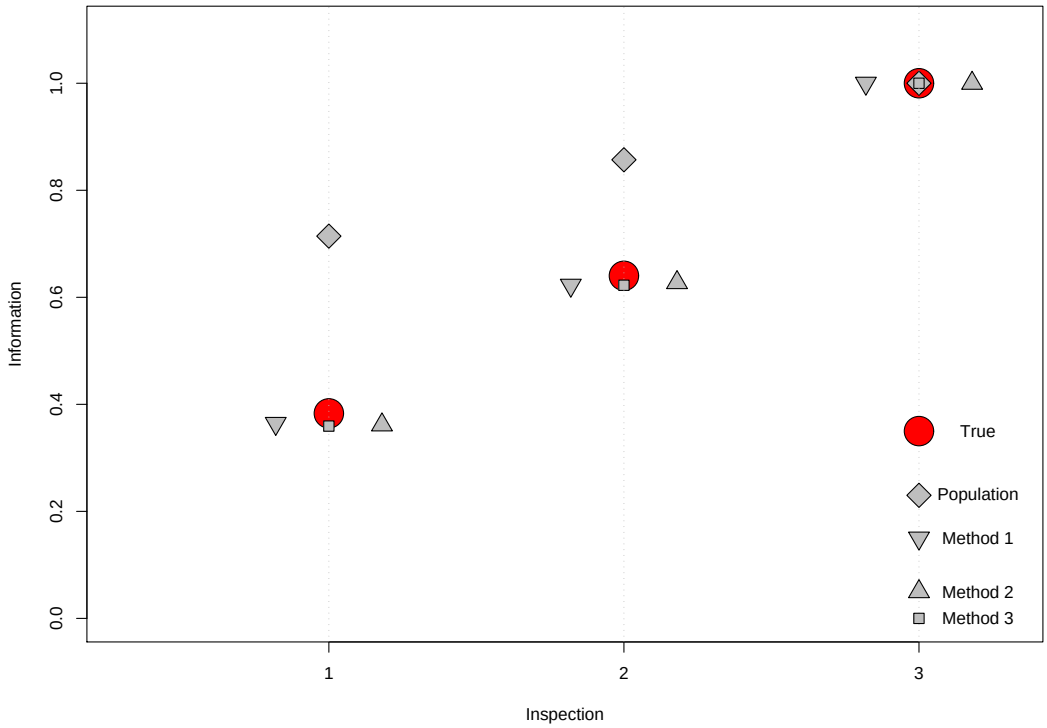


Figure 4.2b: *Information estimation with three interim inspections: BCD(3/4) and nonlinear trend in the data,* $L = 3$*;* $r_1 = 250$ *and* $n_{11} = 126$*,* $r_2 = 300$ *and* $n_{12} = 148$*,* $n = 350$ *and* $n_1 = 174$*.*

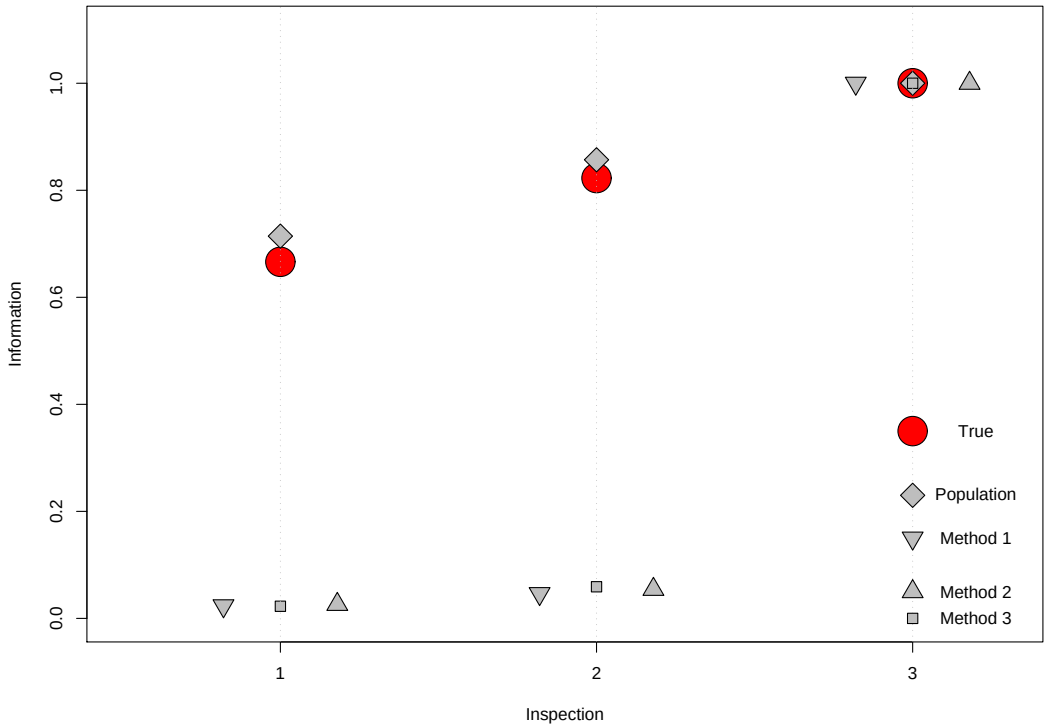


Figure 4.2c: *Information estimation with three interim inspections: BCD(3/4) and linear trend in the data,* $L = 3$*;* $r_1 = 250$ *and* $n_{11} = 126$*,* $r_2 = 300$ *and* $n_{12} = 148$*,* $n = 350$ *and* $n_1 = 174$*.*

Table 4.1: *Information simulations by method of interpolating the unknown observations. Centered simple scores;* $L = 3$, $n = 350$, $n_1 = 174$. *The information at the final look is always 1.*

Complete Randomization			BCD(3/4)	
Method	Look 1	Look 2	Look 1	Look 2
	$r_1 = 250$	$r_2 = 300$	$r_1 = 250$	$r_2 = 300$
	$n_{11} = 126$	$n_{12} = 148$	$n_{11} = 126$	$n_{12} = 148$
No time trend			No time trend	
True	0.3635	0.6282	0.3759	0.6382
Population	0.7143	0.8570	0.7143	0.8570
Method 1	0.3644	0.6286	0.3791	0.6380
Method 2	0.3646	0.6286	0.3743	0.6417
Method 3	0.3666	0.6293	0.3735	0.6352
Nonlinear time trend			Nonlinear time trend	
True	0.3974	0.6488	0.3831	0.6402
Population	0.7143	0.8570	0.7143	0.8570
Method 1	0.3638	0.6229	0.3643	0.6225
Method 2	0.3649	0.6229	0.3616	0.6275
Method 3	0.3662	0.6237	0.3594	0.6225
Linear time trend			Linear time trend	
True	0.9713	0.9893	0.6663	0.8228
Population	0.7143	0.8570	0.7143	0.8570
Method 1	0.3644	0.5559	0.0235	0.0463
Method 2	0.3654	0.5665	0.0262	0.0543
Method 3	0.3661	0.5776	0.0227	0.0591

1. Estimate the boundary for an α–level sequential test with L interim looks based on the algorithm described in Sections 4.1 and 4.2, by sampling $N_c = 10,000$ sequence sequentially from $\bigcap_{i=1}^{l}\{N_1(r_i) = n_{1i}\}$, for each $l = 1, 2, ..., L$. One can use the quantile estimator by Chen and Lazar (2010) or simply use the sample quantile estimator in (4.4). We use (4.4). Denote this boundary by $\hat{d}_1, ..., \hat{d}_L$.

2. Generate at least $N_c = 2500$ sequences from $\bigcap_{i=1}^{L}\{N_1(r_i) = n_{1i}\}$.

3. For each sequence in Step 2, evaluate the linear rank statistic at each inspection l and compare this set of statistics to the estimated boundary $\hat{d}_1, ..., \hat{d}_L$. A rejection occurs if any of the L linear rank statistics exceeds the corresponding boundary. Compute the proportions of rejections. This proportion is an estimate of the probability of type I error.

4. Repeat Steps 2 and 3 $\mathcal{L}$ times to obtain the approximate distribution of the probability of type I error. Obtain the average of these proportions and denote this average by $\hat{\alpha}$.

To obtain the power for an α–level test, we use the following algorithm:

1. Generate at least $N_c = 2500$ sequences from the set satisfying $\bigcap_{i=1}^{L}\{N_1(r_i) = n_{1i}\}$.

2. For each sequence, apply the location–shift to $\boldsymbol{z}$ to create $\boldsymbol{z}_1$. If the jth assignment is to treatment 1, $j = 1, ..., n$, $z_{1j} = z_j + \delta$, and $z_{1j} = z_j$, otherwise.

3. For each sequence, evaluate the linear rank statistic at each inspection l and compare this set of statistics to $\hat{d}_1, ..., \hat{d}_L$. A rejection occurs if any of the L linear rank statistics exceeds this boundary.

Table 4.2: *Mean (S.D.) of estimated α and power for an $\alpha = 0.05$ upper tail sequential test over $\mathcal{L} = 1000$ replications, $N_c = 2500$. Bootstrapping the unknown data sequentially. BCD(3/4). Centered simple scores.*

Look l	r_l	n_{1l}	t_l	α_l^*	$\widehat{d}_l^{**}$	$\widehat{\alpha}$	δ	power
Look 1	250	126	0.3617	0.0011	1709		0.35	
Look 2	300	148	0.6248	0.0121	1688		0.35	
Look 3	350	174	1	0.0373	1501	0.0495 (0.0043)	0.35	0.9709 (0.0035)

* $\alpha_l = \frac{\alpha^*(t_l) - \alpha^*(t_{l-1})}{1 - \alpha^*(t_{l-1})}$; ** Estimated using (4.4) with 10,000 sequences.

4. Compute the proportions of rejections. This proportion is an estimate of the power.

5. Repeat Steps 1–4 $\mathcal{L}$ times to obtain the approximate distribution of power of the upper-tail, α–level test and δ shift. Obtain the average of these proportions to estimate power.

We generate a sample of $n = 350$ observations from $N(1, 0.9)$ and simulate treatment assignments from a BCD($p = 3/4$). We plan $L = 3$ interim looks: at $r_1 = 250$, $r_2 = 300$ and $r_3 = 350$. The observed number assigned to treatment 1 at each look was $n_{11} = 126$, $n_{12} = 128$ and $n_{13} = 174$. Table 4.2 gives the estimated error rates and standard deviations over 1000 replications for this sequential conditional test. The probability of type I error is well preserved. In Table 4.3, we use complete randomization on the same data. The critical values differ and for this example the power estimates are similar.

Table 4.3: *Mean (S.D.) of estimated* α *and power for an* $\alpha = 0.05$ *upper tail sequential test over* $\mathcal{L} = 1000$ *replications,* $N_c = 2500$. *Bootstrapping the unknown data sequentially. Complete randomization. Centered simple scores.*

Look l	r_l	n_{1l}	t_l	α_l^*	$\widehat{d}_l^{**}$	$\widehat{\alpha}$	δ	power
Look 1	250	126	0.3668	0.0012	1742		0.35	
Look 2	300	148	0.6287	0.0122	1669		0.35	
Look 3	350	174	1	0.0371	1587	0.0474 (0.0021)	0.35	0.9669 (0.0018)

* $\alpha_l = \frac{\alpha^*(t_l) - \alpha^*(t_{l-1})}{1 - \alpha^*(t_{l-1})}$; **2 Estimated using (4.4) with 10,000 sequences.

Chapter 5: Conclusions

We have derived the exact conditional distribution of $N_1(n)$, given $N_1(j)$, for Efron's BCD(p) using combinatoric arguments. This result enables us to generate randomization sequences directly from the conditional reference set, which provides a simple and straightforward method for approximating randomization tests. A Monte Carlo sample size of 2500 sequences is adequate for estimating p–values regardless of the values of n, p, or the score functions. This technique extends to stratified test.

We have also shown how to implement the sequential monitoring of conditional randomization tests following complete randomization and the BCD(p) and how to approximate such tests. The approximation of these tests was facilitated by the idea of sampling from the conditional reference set and the Markovian property of $N_1(j)$. In the sequential monitoring case, we estimate the quantiles associated with critical regions rather than test p–values. A difficulty with quantile estimation is the ambiguity related to the requisite sample size. In our simulations, using 10,000 sequences to approximate a population quantile with the sample quantile estimator at each interim look worked well. A quantile estimator for discrete distributions, that comes with a sample size suggestion of 500 to 10,000, exists (Chen and Lazar, 2010) but it remains unclear if, for our purposes, that estimator would perform any better than the simple sample quantile estimator. We have also derived expressions for the exact computation of the conditional variance of the randomization–based linear rank statistic that we use in the definition of the randomization–based information fraction. When approximating the information fraction as the ratio of the interim variance to the total variance of the test statistic, one must interpolate the unknown

data. In our simulations, we found that if the data have a linear trend, interpolation by bootstrapping is extremely conservative. It does not work as well as in the case when there are no trends in the data. Time trends in the data is a special case when randomization tests are known to give different p–values than population–based tests. For this reason, the estimation of information in the presence of time trends requires more attention.

The class of generalized biased coin designs (Wei, 1978) does not have a known form for the exact conditional distribution, and therefore we cannot apply this technique. Mehta, Patel and Senchaudhuri (1988b) are able to estimate conditional randomization tests for this class of designs by using importance sampling. The conditional linear rank test is asymptotically normal under some score functions for the generalized biased coin design (Smythe, 1988), which provides perhaps the simplest method of all for computing tests, when there are large samples. For the sequential monitoring of conditional tests using the GBCD with one interim look, Zhang and Rosenberger (2006) derive the joint asymptotic distribution of the interim and the final test statistics, which allows for an asymptotic test. In the sequential monitoring case with two or more interim inspections, an importance sampling technique may be needed for the approximation of tests following the GBCD, unless exact properties of the GBCD similar to those of the BCD(p) are obtained. Under Efron's BCD(p), the conditional randomization test is not asymptotically normal and therefore our technique appears to provide a reasonable method that can be programmed quite easily using Theorem 2.1. This technique also extends naturally to the sequential monitoring of conditional randomization tests.

The critical values $d_1, d_2, ..., d_L$, estimated in Section 4.2, provide a boundary that detects efficacy, or evidence in favor of the alternative, at the nominal probability of type I error. Similarly, a boundary that detects futility, or evidence against the

alternative, at the nominal probability of type II error can be estimated as well. Furthermore, two boundaries that test for efficacy and futility, while simultaneously maintaining the type I and type II error probabilities, can also be estimated. In the future, we plan to show how to estimate such boundaries for conditional randomization tests. In addition, we hope to formalize the sequential monitoring of permutation–based linear rank tests under the population model. Such a formulation consists of arguments about conditioning on sufficient statistics rather than conditioning on $N_1(r_l)$.

List of Symbols

Bibliography

[1] Baldi Antognini, A. (2008). "A Theoretical Analysis of the Power of Biased Coin Designs," *Journal of Statistical Planning and Inference*, 138, 1792–1798.

[2] Chen, J. and Lazar, N. (2010). "Quantile Estimation for Discrete Data via Empirical Likelihood," *Journal of Nonparametric Statistics*, 22, 237-255.

[3] Cytel Inc. (2007). *StatXact 8: Statistical Software for Exact Nonparametric Inference, Volume 2*. Cambridge: Cytel Inc.

[4] Cox, D. R. (1982). "A Remark on Randomization in Clinical Trials," *Utilitas Mathematica*, 21A, 245–252.

[5] Edgington, E.S. and Onghena, P. (2007). *Randomization Tests*. Chapman & Hall/CRC, Boca Raton.

[6] Efron, B. (1971). "Forcing a Sequential Experiment to Be Balanced," *Biometrika*, 58, 403–417.

[7] Flyer, P. A. (1998). "A Comparison of Conditional and Unconditional Randomization Tests for Highly Stratified Data," *Biometrics*, 54, 1551–1559.

[8] Good, P. (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer, New York.

[9] Hollander, M. and Peña, E. (1988). "Nonparametric Tests Under Restricted Treatment-Assignment Rules," *Journal of the American Statistical Association*, 83, 1141–1151.

[10] Kalish, L. A. and Begg, C. B. (1987). "The Impact of Treatment Allocation Procedures on Nominal Significance Levels and Bias," *Controlled Clinical Trials*, 8, 121–135.

[11] Kempthorne, O. (1986). "Randomization–II." In *Encyclopedia of Statistical Sciences, Volume 7.* S. Koltz and N. L. Johnson, Eds., 519–524. John Wiley & Sons, New York.

[12] Lan, K. K. G. and DeMets, D. L. (1983). "Discrete Sequential Boundaries for Clinical trials," *Biometrika*, 70, 659–663.

[13] Lehmann, E. L. (1986). *Testing Statistical Hypotheses.* John Wiley & Sons, New York.

[14] Markaryan, T. and Rosenberger, W. F. (2010). "Exact Properties of Efron's Biased Coin Randomization Procedure," *Annals of Statistics*, 38, 1546–1567.

[15] Mehta, C. R., Patel, N. R., and Wei, L. J. (1988a). "Constructing Exact Significance Tests With Restricted Randomization Rules," *Biometrika*, 75, 295–302.

[16] Mehta, C. R., Patel, N. R., and Senchaudhuri, P. (1988b). "Importance Sampling for Estimating Exact Probabilities in Permutational Inference," *Journal of the American Statistical Association*, 83, 888–1005.

[17] O'Brien, P. C. and Fleming, T. R. (1979). "A Multiple Testing Procedure for Clinical Trials," *Biometrics*, 35, 549–556.

[18] Pesarin, F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. Chichester: New York.

[19] Prentice, R. L. (1978). "Linear Rank Tests with Right–Censored Data," *Biometrika*, 65, 167–179.

[20] Rosenberger, W. F. and Lachin, J. M. (2002). *Randomization in Clinical Trials: Theory and Practice*. New York: Wiley.

[21] Shao J. (2003). *Mathematical Statistics*. Springer, New York.

[22] Smith, R. (1984). "Properties of Biased Coin Designs in Sequential Clinical Trials," *Annals of Statistics*, 12, 1018–1034.

[23] Smythe, R. T. (1988). "Conditional Inference for Restricted Randomization Designs," *Annals of Statistics*, 16, 1155–1161.

[24] Smythe, R. T. and Wei, L. J. (1983). "Significance Tests With Restricted Randomization Design," *Biometrika*, 70, 496–500.

[25] Wei, L. J. (1978). "The Adaptive Biased Coin Design for Sequential Experiments," *Annals of Statistics*, 6, 92–100.

[26] Wei, L. J., Smythe, R. T., and Smith, R. L. (1986). "K–Treatment Comparisons With Restricted Randomization Rules in Clinical Trials," *Annals of Statistics*, 14, 265–274.

[27] Zhang, L. and Rosenberger, W. F. (2010). "Adaptive Randomization in Clinical Trials." In *Design and Analysis of Experiments, Volume III.* K. Hinkelmann, Eds., in press.

[28] Zhang, Y. and Rosenberger, W. F. (2008). "Sequential Monitoring of Randomization Tests: Generalized Biased Coin Designs," *Sequential Analysis*, 27, 234–253.

Curriculum Vitae

Victoria Plamadeala graduated from the Academy of Economic Studies of Moldova in 2002 with an undergraduate degree in international business and economics. She studied economics at the University of Central Florida, where she first learnt of statistical inference. While being a graduate student in statistics at George Mason University, she worked in the Statistical Consulting Center and was a summer intern at The Dupuy Institute in Annandale, VA analyzing insurgency data.