Fast permutation inference for randomized experiments with binary outcomes

P. M. Aronow Haoge Chang Patrick Lopatto*

Abstract

Consider a randomized experiment with a binary outcome. In this setting, confidence intervals for the average causal effect can be determined through a series of permutation tests. This approach requires minimal assumptions and is valid for all sample sizes, as it does not rely on large-sample approximations such as the central limit theorem. We show that these confidence intervals can be found in $O(n \log n)$ permutation tests when the experiment is balanced. Further, we demonstrate how to construct confidence intervals using Monte Carlo permutation tests, instead of more computationally demanding exact tests, while retaining the guarantee that the intervals contain the true parameter at the given rate. Our results render permutation inference applicable to experiments far larger than those accessible by previous methods.

*P.L. is partially supported by NSF postdoctoral fellowship DMS-220289.

Contents

1 Introduction

There is vast literature on causal inference in randomized experiments with binary outcomes (see, for example, [IR15] and the references therein). While the assumption of binary outcomes may seem restrictive, this setting nonetheless includes a large class of important examples, such as experiments probing the causal effect of a vaccine for a new disease, a design change in an online advertisement, or a reminder about the date of an upcoming election. In each case, the outcome for each subject – whether they contract an infection, click a banner, or cast a ballot, respectively – is binary.

Many traditional approaches to analyzing experiments with binary outcomes are based on a binomial model for the outcome distribution. Examples include Wald confidence intervals [Was04, Chapter 10], which are based on a normal approximation valid in large samples, as well as more sophisticated interval estimators that do not make such approximations and are effective for all sample sizes [BB19, SS80]. However, the assumptions underlying the binomial model are highly problematic in typical experimental designs. This was noted by Robins in [Rob88], whose discussion we now briefly summarize.

Consider an experiment with n subjects, where m subjects are assigned to treatment and the remaining $n - m$ are assigned to control, and the random assignment is such that all possible configurations are equally likely. Using the observed outcomes, we wish to construct a confidence interval $\mathcal{I}_\alpha$ that contains the average causal effect of the treatment with probability $1 - \alpha$ for some given $\alpha \in (0, 1)$. (A conventional choice is $\alpha = 0.05$, leading to a 95% confidence interval.) Let the outcome space be $\{0, 1\}$. Robins identifies two modeling assumptions under which binomial intervals will cover the true average causal effect at the nominal rate.

1. If a subject is exposed to treatment, their outcome may be modeled as a Bernoulli random variable, and the treatment outcomes across subjects are independent with common mean p_1. Similarly, the control outcomes are independent Bernoulli random variables with common mean p_2. Define the average causal effect of treatment as $p_1 - p_2$.

2. The subjects in the study are drawn uniformly at random from some near-infinite superpopulation. We let p_1 be the proportion of subjects in the superpopulation who whose outcome would be 1 under treatment, and similarly let p_2 be the proportion whose outcome would be 1 under control. Again define the average causal effect of treatment as $p_1 - p_2$.

Robins notes that Assumption (1) is untenable in most contexts, since it does not account for between-subject variation. For example, disease risk may vary among subjects in a clinical trial according to pre-existing medical conditions and demographic characteristics. Further, Assumption (2) is almost always false, since typical subject recruitment strategies do not result in a uniform sample from a well-defined superpopulation. Consider, for example, a clinical trial where subjects are recruited through newspaper advertisements. It is generally implausible that all members of the target population will read and respond to such advertisements at identical rates.

Instead of the assumptions above, we adopt the more flexible potential outcomes framework for studying causality, which originated in work of Neyman and has subsequently been developed by many researchers (see [IR15] for details). Suppose we have a group of n subjects, and label them arbitrarily from 1 to n. Our fundamental assumption is that the observed outcome for subject i depends only on the index i and whether that subject is assigned to treatment or control. This is known as the *stable unit treatment value assumption*, and abbreviated SUTVA. In the particular, SUTVA rules out between-subject interaction effects. We let $\boldsymbol{w}_i = \big(w_i(1), w_i(2)\big)$ with $w_i(1), w_i(2) \in \{0, 1\}^2$ denote the *potential outcomes* for the i-th subject, where $w_i(1)$ is the outcome that would be observed if the subject were assigned to treatment, and $w_i(2)$ is the outcome that would be observed if the subject were assigned to control. In the experiment, only *one* of these potential outcomes is observed; the other remains unknown. We aim to estimate the *sample average treatment effect*, defined as

$$\mathrm{SATE} = \frac{1}{n}\sum_{j=1}^{n}\big(w_j(1) - w_j(2)\big). \tag{1.1}$$

The quantity (1.1) is defined using only the subjects in the experiment, and says nothing about the causal effect in any larger population. Such a generalization would require additional assumptions. However, precise estimates of (1.1) are still broadly useful for studying causal claims. For example, if a company asserts that their latest drug greatly reduces the chance of stroke in elderly people, and a study of the average causal effect in a group of such people produces a precise estimate centered at 0, it is reasonable to reject the company's claim.

Since we view the potential outcomes of the subjects as fixed, the only randomness in the experiment comes from the assignment of subjects to treatment and control. We retain our previous setup where m subjects are assigned to treatment, the remainder are assigned to control, and all such assignments are equally likely. Under this design, Robins shows that the binomial Wald interval typically does not cover the true parameter (1.1) at the nominal rate [Rob88]. In fact, it may dramatically undercover, and the probability that the Wald interval contains (1.1) tends to zero for certain configurations of potential outcomes as n grows large. Robins also gives a confidence interval for (1.1) that relies on a large-sample normal approximation. While it generally has better coverage than binomial intervals when n is large, there is no quantitative estimate available for its accuracy. Further, it behaves poorly in smaller samples, and in large samples for certain configurations of the potential outcomes.

Recently, it was realized by Rigdon and Hudgens that *exact* confidence intervals, which cover at the nominal rate regardless of the sample size or potential outcomes, could be constructed by inverting permutation tests [RH15]. While the observation that one can use permutation inference to test the compatibility of the observed data with a *single* set of potential outcomes was essentially known to Fisher in the 1930s [IR15, Chapter 5], Rigdon and Hudgens propose testing the compatibility of *every* possible set of potential outcomes with the observed data. From this, one can collect the sample average treatment effects of all compatible sets of potential outcomes to form a confidence interval for (1.1). We provide complete details regarding this method below, in Section 2.

While the permutation approach eliminates any reliance on uncontrolled large-sample approximations, it requires significant computational resources. The bulk of the computation arises from two sources, the number of permutation tests required to construct the interval, and the effort required to conduct each permutation test. Regarding the first contribution, in the original algorithm of Rigdon and Hudgens, $O(n^4)$ permutation tests are required in the worst case. Later, Li and Ding showed that the same intervals could be constructed in $O(n^2)$ permutation tests in the balanced case, where an equal number of subjects are assigned to treatment and control (that is, $n = 2m$), and conjectured that their algorithm also reproduces the Rigdon–Hudgens intervals in unbalanced trials [LD16]. (They showed that the interval their algorithm returns always contains the corresponding Rigdon–Hudgens interval, but were unable to exclude the possibility that theirs are strictly larger for unbalanced designs.)

We next consider the computational burden of performing a single permutation test. An exact test enumerates all $\binom{n}{m}$ possible assignments of subjects, which grows exponentially in n and quickly becomes computationally infeasible. Rigdon and Hudgens therefore recommend performing approximate permutation tests, where the permutation p-value is approximated by Monte Carlo simulation with a fixed number of samples K. However, such intervals may fail to cover at the nominal rate due to Monte Carlo error, and there has been no discussion in the literature regarding how large K should be taken so that the resulting intervals have good coverage.

We make two contributions, corresponding to the two sources of computational effort identified previously. We focus on the case of balanced experiments, where the treatment and control groups have equal size, which is common in practice (as it maximizes the statistical efficiency of the experiment relative to unbalanced designs). First, we show that the exact confidence intervals from [RH15] can be constructed in $O(n \log n)$ permutation tests, improving on the $O(n^2)$ tests required by [LD16]. We also extend our analysis to experiments with missing data, a common statistical problem that has not been addressed previously in this context. Second, we show how to construct confidence intervals using approximate permutation tests that are guaranteed to cover the sample average treatment effect with probability $1-\alpha$ for any $\alpha > 0$. We further show that by increasing K, these intervals will approximate those of [RH15] arbitrarily well, and quantify how K must scale in order to guarantee any desired coverage rate and approximation accuracy as n increases. Together, our contributions permit permutation inference to be applied to experiments far larger than those accessible by previous methods.

Many interesting questions remain. First, can the confidence intervals we produce in $O(n \log n)$ permutation tests be constructed in asymptotically fewer tests? Second, to what extent is computationally efficient randomization inference possible for unbalanced trials? Third, is efficient randomization inference possible in the case of a general bounded discrete outcome, or for a continuous outcome supported on a bounded interval? Finally, does there exist a practical algorithm for prospective sample size calculations? Researchers often wonder how many experimental subjects they need to detect an effect with high probability, assuming the effect is larger than some given threshold. If too few subjects are used, the experiment is uninformative, and if too many subjects are used, the experiment is potentially wasteful.

We begin in Section 2 by introducing our notation and reviewing previous work in detail. In Section 3, we present basic properties of the permutation intervals constructed in [RH15]. In Section 4, we show how to construct these intervals in $O(n \log n)$ permutation tests, and in Section 5, we consider missing data. In Section 6, we construct exact intervals using Monte Carlo permutation tests. The remaining sections contain proofs of the assertions in the previous sections.

2 Preliminaries

2.1 Notation

Let $\mathbb{Z}_{>0}$ denote the positive integers. We consider a group of $n \in \mathbb{Z}_{>0}$ subjects who undergo an experiment with a binary outcome. We always suppose that the experiment is balanced, so that $n = 2m$ for some $m \in \mathbb{Z}_{>0}$, m subjects are assigned to treatment, and the remaining m subjects are assigned to control.

A *potential outcome table* is any vector

$$\boldsymbol{w} \in \big\{(0,0),(0,1),(1,0),(1,1)\big\}^n. \tag{2.1}$$

Letting $\boldsymbol{w}_i = \big(w_i(1), w_i(2)\big)$ for $i \in [\![1,n]\!]$, the coordinate $w_i(1)$ denotes the outcome for the i-th subject under treatment, and $w_i(2)$ denotes the outcome under control. We use $\boldsymbol{y}$ to denote the potential outcome table for the subjects in the experiment, and $\boldsymbol{w}$ to denote a potential outcome table for a hypothetical group of subjects. Given some table $\boldsymbol{w}$, we let the corresponding count vector $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{w}) \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$ be defined by

$$\boldsymbol{v} = (v_{11}, v_{10}, v_{01}, v_{00}), \qquad v_{ab} = \sum_{j=1}^{n} \mathbb{1}\{\boldsymbol{w}_j = (a, b)\}. \tag{2.2}$$

We now introduce notation for the randomization of the subjects. Consider the random vector

$$\boldsymbol{Z} = (Z_1, Z_1, \dots, Z_n) \in \{0, 1\}^n \tag{2.3}$$

whose distribution is uniform over the set

$$\mathcal{Z}(n) = \left\{ \boldsymbol{z} \in \{0, 1\}^n : \sum_{i=1}^{n} z_i = m \right\}. \tag{2.4}$$

Here, the indices i such that $Z_i = 1$ represent the subjects assigned to treatment. The vector of observed experimental outcomes $\boldsymbol{Y} = (Y_j)_{j=1}^n$ is then given by

$$Y_j = Z_j y_j(1) + (1 - Z_j) y_j(2). \tag{2.5}$$

We say that a table $\boldsymbol{w}$ is *possible* given the observed data $\boldsymbol{Y}$ if for every $j \in [\![1, n]\!]$, we have $\boldsymbol{w}_j(1) = Y_j$ if $Z_j = 1$, and $\boldsymbol{w}_j(2) = Y_j$ if $Z_j = 0$. We say that a count vector $\boldsymbol{v}$ is possible if it arises from some possible table $\boldsymbol{w}$.

The sample average treatment effect for a potential outcome table is defined by

$$\tau(\boldsymbol{w}) = \frac{1}{n} \sum_{j=1}^{n} \big(w_j(1) - w_j(2) \big). \tag{2.6}$$

We define the Neyman estimator of $\tau(\boldsymbol{y})$ by

$$T(\boldsymbol{Y}, \boldsymbol{Z}) = \sum_{j=1}^{n} \frac{Z_j Y_j}{m} - \sum_{j=1}^{n} \frac{(1 - Z_j) Y_j}{n - m}, \tag{2.7}$$

and note that $\mathbb{E}\big[T(\boldsymbol{Y}, \boldsymbol{Z})\big] = \tau(\boldsymbol{y})$, where the expectation is taken over the variable $\boldsymbol{Z}$. This estimator depends on $(\boldsymbol{Y}, \boldsymbol{Z})$ only through the observed count vector $\boldsymbol{n} \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$ defined by

$$\boldsymbol{n} = (n_{11}, n_{10}, n_{01}, n_{00}), \qquad n_{zy} = \sum_{j=1}^{n} \mathbb{1}\{Z_j = z, Y_j = y\}, \tag{2.8}$$

for $z \in \{0, 1\}$ and $y \in \{0, 1\}$. For example, n_{10} is the number of subjects in the treatment group with observed outcome 0. We overload the notation slightly and write $T(\boldsymbol{n})$ for the value of the Neyman estimator $T(\boldsymbol{Y}, \boldsymbol{Z})$, if $\boldsymbol{n}$ is the observed count vector for $(\boldsymbol{Y}, \boldsymbol{Z})$. We will do this without further comment for all other functions of $(\boldsymbol{Z}, \boldsymbol{Y})$ that depend only on the associated vector $\boldsymbol{n}$, and similarly for functions of tables $\boldsymbol{w}$ that depend only on the associated count vector $\boldsymbol{v}$.

Throughout the paper, $[\![a, b]\!]$ denotes the set $\{k \in \mathbb{Z} : a \leq k \leq b\}$. We write log for the natural logarithm. When we use the base 2 logarithm, it is always denoted by $\log_2$. Additionally, we often abbreviate probability mass function as pmf. Further notation, used only in the proofs, is introduced in Section 7.

2.2 Previous Work

2.2.1 Rigdon and Hudgens

We begin by recalling the confidence interval procedure proposed in [RH15]. As $\boldsymbol{w}$ ranges over all tables, $\tau(\boldsymbol{w})$ takes every value in the set

$$\mathcal{S}(n) = \left\{ -\frac{n}{n}, -\frac{n-1}{n}, \dots, \frac{0}{n}, \dots, \frac{n-1}{n}, \frac{n}{n} \right\}, \tag{2.9}$$

which has $2n+1$ elements. After the experiment is completed, Rigdon and Hudgens note that the set of $\tau(\boldsymbol{w})$ for $\boldsymbol{w}$ that are possible given the observed data is further restricted to

$$\mathcal{C}(\boldsymbol{Y}, \boldsymbol{Z}) = \left\{ \frac{1}{n} \left(\sum_{j=1}^{n} Y_j(2Z_j - 1) - m \right), \dots, \frac{1}{n} \left(\sum_{j=1}^{n} Y_j(2Z_j - 1) - m + n \right) \right\}, \tag{2.10}$$

which has $n+1$ elements. We observe that $\mathcal{C}$ depends on $(\boldsymbol{Y}, \boldsymbol{Z})$ only through the associated observed count vector $\boldsymbol{n}$.

We now consider some $\alpha \in (0,1)$ and a realization of $(\boldsymbol{Y}, \boldsymbol{Z})$. We will construct a confidence interval for τ at level $1-\alpha$ (that is, one that contains the true parameter with probability at least $1-\alpha$). We first describe a permutation test for the sharp null hypothesis that $\boldsymbol{y} = \boldsymbol{w}$ for some given $\boldsymbol{w}$, then explain how to form a confidence interval from a series of these tests.

Definition 2.1. Let $\widetilde{\boldsymbol{Z}}$ be a random vector independent from $\boldsymbol{Z}$ with the same distribution, and let $\widetilde{\boldsymbol{Y}}$ be the vector with entries

$$\widetilde{Y}_j = \widetilde{Z}_j w_j(1) + (1 - \widetilde{Z}_j) w_2(2) \tag{2.11}$$

We say that $\boldsymbol{w}$ is compatible with data $(\boldsymbol{Y}, \boldsymbol{Z})$ at level $1-\alpha$ if[1]

$$p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) = \mathbb{P}\left(\left| T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) - \tau(\boldsymbol{w}) \right| \geq \left| T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau(\boldsymbol{w}) \right| \right) \geq \alpha. \tag{2.12}$$

We say that $\boldsymbol{w}$ is compatible with an observed count vector $\boldsymbol{n}$ if (2.12) holds with $T(\boldsymbol{Y}, \boldsymbol{Z})$ replaced by $T(\boldsymbol{n})$, and define $p(\boldsymbol{w}, \boldsymbol{n})$ similarly.

Let $\boldsymbol{v}$ denote the count vector associated to $\boldsymbol{w}$, as defined in (2.2), and let $\boldsymbol{n}$ be as in (2.8). Then it is straightforward to see that $p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) = p(\boldsymbol{v}, \boldsymbol{n})$, so that the p-value depends on $(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z})$ only through the count vectors $\boldsymbol{n}$ and $\boldsymbol{v}$.

Algorithm 2.2. A *permutation test* takes as inputs $\alpha \in (0,1)$, a potential outcome table $\boldsymbol{w}$, and a vector $\boldsymbol{n} \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$. It returns a binary decision, accept or reject, as follows. The algorithm computes $p(\boldsymbol{w}, \boldsymbol{n})$ by direct enumeration of all $\binom{n}{m}$ elements of the set $\mathcal{Z}(n)$ (defined in (2.4)), over which $\widetilde{\boldsymbol{Z}}$ is uniformly distributed. If $p(\boldsymbol{w}, \boldsymbol{n}) \geq \alpha$, it accepts. Otherwise, it rejects.

We now present the confidence interval procedure from [RH15].

[1] In fact, the inequality $\geq \alpha$ can be replaced with the inequality $> \alpha$ in (2.12), which creates a stricter threshold for acceptance. This change leads to potentially shorter confidence intervals that still cover at the nominal rate, which can be seen by inspecting the proof of Proposition 3.2. Further, all of our results are still valid under the stricter definition. The definition (2.1) was used in the previous works [RH15, LD16], and we retain it for consistency.

Algorithm 2.3. The following algorithm takes as input $\alpha \in (0,1)$ and a vector $\boldsymbol{n} \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$. It returns an interval $\mathcal{I}_\alpha(\boldsymbol{n}) = \big[U_\alpha(\boldsymbol{n}), L_\alpha(\boldsymbol{n})\big]$. We first give the steps for finding the upper limit $U_\alpha(\boldsymbol{n})$.

1. Let τ_0 equal the largest element of $\mathcal{C}(\boldsymbol{n})$ (defined in (2.10)).
2. For each possible count vector $\boldsymbol{v}$ such that $\tau(\boldsymbol{v}) = \tau_0$, apply Algorithm 2.2. If all possible count vectors are rejected, then reject τ_0. Otherwise, accept τ_0.
3. If τ_0 is accepted, then set $U(\boldsymbol{n}) = \tau_0$. Otherwise, repeat these steps with the value $\tau_0 - n^{-1}$ until some value is accepted, and set $U(\boldsymbol{n})$ equal to this value.

The lower limit $L(\boldsymbol{n})$ is found analogously, starting from the smallest element of $\mathcal{C}$ and iterating upward in increments of n^{-1}.

Observe that the third step of the algorithm does not say what to do in the event that all elements of $\mathcal{C}$ are rejected. While [RH15] do not discuss this point explicitly, it is simple to demonstrate the algorithm will always terminate by returning values for U_α and L_α such $L_\alpha \leq \tau(\boldsymbol{y}) \leq U_\alpha$. This is shown in Lemma 3.1 below.

As noted previously, any observed data $(\boldsymbol{Y}, \boldsymbol{Z})$ gives rise to an observed count vector $\boldsymbol{n}$. Then using Algorithm 2.3, we define

$$\mathcal{I}_\alpha(\boldsymbol{Y}, \boldsymbol{Z}) = \big[L_\alpha(\boldsymbol{Y}, \boldsymbol{Z}), U_\alpha(\boldsymbol{Y}, \boldsymbol{Z})\big] \tag{2.13}$$

by setting $\mathcal{I}_\alpha(\boldsymbol{Y}, \boldsymbol{Z}) = \mathcal{I}_\alpha(\boldsymbol{n})$. Rigdon and Hudgens observe that since there are $O(n^4)$ possible count vectors $\boldsymbol{v}$ associated to possible tables $\boldsymbol{w}$, the interval $\mathcal{I}_\alpha$ can be found in $O(n^4)$ permutation tests [RH15]. In fact, since the first three coordinates of $\boldsymbol{v}$ determine the fourth (as they sum to n), only $O(n^3)$ permutation tests are needed. Further, this algorithm does not require our standing hypothesis that the treatment and control groups have an equal number of subjects, and may be applied to arbitrary unbalanced trials.

2.2.2 Li and Ding

As noted in the introduction, Li and Ding have proposed an algorithm that returns $I_\alpha(\boldsymbol{n})$ in $O(n^2)$ permutation tests for balanced trials [LD16]. We do not give the details here, since they are not relevant to our work. However, we will make use of the following lemmas from [LD16]. The first provides a necessary and sufficient condition for checking whether a potential outcome table is possible given the observed data.

Lemma 2.4 ([LD16, Theorem 1])**.** *A potential outcome table $\boldsymbol{w}$ with count vector $\boldsymbol{v}$ is possible given observed data $\boldsymbol{n}$ if and only if*

$$\begin{aligned} &\max(0, n_{11} - v_{10}, v_{11} - n_{01}, v_{11} + v_{01} - n_{10} - n_{01}) \\ &\leq \min(v_{11}, n_{11}, v_{11} + v_{01} - n_{01}, n - v_{10} - n_{01} - n_{10}). \end{aligned} \tag{2.14}$$

The second lemma shows that in the balanced case, the confidence set that we would obtain if we tested *every* possible value of τ_0 is indeed an interval.

Lemma 2.5 ([LD16, Theorem A.4])**.** *Fix $\alpha \in (0,1)$ and observed data $\boldsymbol{y}$ and $\boldsymbol{Z}$. For every $\tau_0 \in \mathcal{I}_\alpha(\boldsymbol{y}, \boldsymbol{Z}) \cap \mathcal{C}$, there exists a possible potential table $\boldsymbol{w}$ such that $\tau(\boldsymbol{w}) = \tau_0$ and $p(\boldsymbol{w}) \geq \alpha$.*

3 Basic Properties

We now note three basic properties of the confidence intervals $\mathcal{I}_\alpha$. While the first two have been mentioned informally in previous works, they have not been proved, and we feel there is value in giving precise statements and justifications. The third is new. The proofs appear in Section 7. The arguments for all results in this section extend to the unbalanced case, where the treatment group has size $m = cn$ for $c \in (0,1)$, with only minor changes. We omit these extensions for brevity.

Our first lemma implies that the interval $I_\alpha(\boldsymbol{n})$ always contains the estimate $T(\boldsymbol{n})$ of $\tau(y)$.

Lemma 3.1. *For any* $\boldsymbol{n} \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$*, we have*

$$L_\alpha(\boldsymbol{n}) \leq T(\boldsymbol{n}) \leq U_\alpha(\boldsymbol{n}). \tag{3.1}$$

The next proposition states that the intervals $\mathcal{I}_\alpha$ are exact.

Proposition 3.2. *Fix potential outcomes* $\boldsymbol{y}$ *and* $\alpha \in (0,1)$*. Then*

$$\mathbb{P}\big(\tau(\boldsymbol{y}) \in \mathcal{I}_\alpha(\boldsymbol{Y}, \boldsymbol{Z})\big) \geq 1 - \alpha, \tag{3.2}$$

where the probability is with respect to the variable $\boldsymbol{Z}$*.*

Finally, the following proposition states that the intervals $\mathcal{I}_\alpha$ converge at a $n^{-1/2}$ rate. This rate is characteristic of many confidence interval procedures based on central limit theorem asymptotics. We therefore see that the large n scaling behavior of the length of the interval $\mathcal{I}_\alpha$ is the same as that of the intervals produced by asymptotic methods.[2]

Proposition 3.3. *For any* $\boldsymbol{n} \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$*, we have*

$$\big|I_\alpha(\boldsymbol{n})\big| \leq \sqrt{\frac{32\log(2/\alpha)}{n}}. \tag{3.3}$$

4 Fast Permutation Intervals

We now present our main result on the efficient computation of $\mathcal{I}_\alpha$. Its proof is given in Section 8, along with the proofs of all other results in this section.

Theorem 4.1. *For any* $\alpha \in (0,1)$ *and observed count vector* $\boldsymbol{n}$*, the interval* $\mathcal{I}_\alpha(\boldsymbol{n})$ *can be constructed using at most* $4(n+1)\lfloor \log_2(n+1) + 2 \rfloor$ *permutation tests.*

To prove Theorem 4.1, we exhibit an algorithm that constructs $\mathcal{I}_\alpha(\boldsymbol{n})$ in the required number of permutation tests.[3] We first recall a version of the standard binary search algorithm (see, for example, [Wik22]).

Algorithm 4.2. This *binary search algorithm* takes as input $k_1, k_2 \in \mathbb{Z}_{>0}$ such that $k_2 > k_1$, and a function $f\colon [\![k_1, k_2]\!] \to \{0,1\}$ such that $f(x) = 0$ if $x \leq r$ and $f(x) = 1$ if $x > r$, where $r \in [\![k_1 - 1, k_2]\!]$ is unknown. The algorithm returns r through the following steps.

[2] In Proposition 3.3, our focus is on establishing the correct scaling rate, not obtaining the optimal constant prefactor. This estimate is quite conservative compared to the interval lengths we find empirically through simulation. It should not be used in practice for prospective sample size calculations.

[3] Again, the estimate in Theorem 4.1 is conservative relative to the empirical performance of Algorithm 4.4 below. To streamline the arguments, we did not pursue the sharpest possible bounds.

1. Set $a = k_1$ and $b = k_2$.
2. Set $c = \lfloor (a+b)/2 \rfloor$ and evaluate $f(c)$. If $f(c) = 0$, set $a = c$. If $f(c) = 1$, set $b = c$.
3. Repeat the previous step until $b = a + 1$. In this case, return a if $a > k_1$ and $b < k_2$. If $a = k_1$, evaluate $f(k_1)$ and return $k_1 - 1$ if $f(k_1) = 1$, and k_1 if $f(k_1) = 0$. If $b = k_2$, evaluate $f(k_2)$, and return k_2 if $f(k_2) = 0$ and $k_2 - 1$ if $f(k_2) = 1$.

It is well known that the previous algorithm terminates after at most $\lfloor \log_2(k_2 - k_1 + 1) + 2 \rfloor$ evaluations of f. For completeness, we prove this below (see Lemma 8.1).

We now present an algorithm that constructs $\mathcal{I}_\alpha(\boldsymbol{n})$ (as will be demonstrated in the proof Theorem 4.1). We begin by indicating the intuition motivating it. Our first observation is that the problem of finding the upper bound U_α reduces to finding an a method to determine whether a given value $\tau_0 \in \mathcal{C}$ is compatible with the observed data. If this can be done in $O(n)$ permutation tests, then Algorithm 4.2 applied to the function $f\colon [\![nT(\boldsymbol{n}), n\max(\mathcal{C}(n))]\!] \to \{0,1\}$ defined by permutation testing x/n for compatibility gives a method for determining U_α in a total of $O(n \log n)$ permutation tests. Here, x is the input to f, and we code compatibility as 0 and incompatibility as 1, recall that the domain of f has cardinality $O(n)$, and note that f satisfies the hypotheses of Algorithm 4.2 by Lemma 2.5. Analogous reasoning applies for finding the lower bound L_α.

Consequently, the proof of Theorem 4.1 focuses on showing that given $\tau_0 \in \mathcal{C}$ can be checked for compatibility in $O(n)$ permutation tests. Observe that any possible count vector $\boldsymbol{v} = (v_{11}, v_{10}, v_{01}, v_{00})$ such that $\tau(\boldsymbol{v}) = \tau_0$ satisfies the equations

$$v_{11} + v_{10} + v_{01} + v_{00} = n, \qquad v_{10} - v_{01} = n\tau_0. \tag{4.1}$$

The following proposition allows us to efficiently search the space cut out by (4.1) by exploiting a certain monotonicity property of the permutation p-values defined in (2.12).

Lemma 4.3. *Fix observed data* $\boldsymbol{n}$*, and a count vector* $\boldsymbol{v} = (v_{11}, v_{10}, v_{01}, v_{00})$*. Suppose* $\min(v_{10}, v_{01}) \geq 1$ *and* $\max(v_{10}, v_{01}) \geq 2$*, and set*

$$\boldsymbol{v}' = (v_{11} + 1, v_{10} - 1, v_{01} - 1, v_{00} + 1). \tag{4.2}$$

Then $p(\boldsymbol{v}', \boldsymbol{n}) \geq p(\boldsymbol{n}, \boldsymbol{v})$*.*

The upshot of this lemma is that rejecting a single $\boldsymbol{v}$ is sufficient to show that all potential outcome tables on the line segment given by the translations (4.2), with endpoint $\boldsymbol{v}$, are also incompatible. Heuristically, the transformation in (4.2) causes the distribution of the variable $T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) - \tau(\boldsymbol{w})$ in the definition (2.12) to become more spread out about its mean (zero), increasing the p-value. This heuristic is made precise in the proof of Lemma 4.3.

Given this context, we now present our main algorithm.

Algorithm 4.4. This algorithm takes as input $\alpha \in (0,1)$ and a vector $\boldsymbol{n} \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$. It returns an interval $\left[L_\alpha^{(1)}(\boldsymbol{n}), U_\alpha^{(1)}(\boldsymbol{n})\right]$. We give only the steps for finding $U_\alpha^{(1)}(\boldsymbol{n})$, since finding $L_\alpha^{(1)}(\boldsymbol{n})$ is analogous.

We perform a binary search using (4.2) with a certain function $f\colon [\![nT(\boldsymbol{n}), n\max(\mathcal{C}(n))]\!] \to \{0,1\}$. Given an input x, the function f returns 0 if x/n is compatible with the vector $\boldsymbol{n}$, and 1 otherwise, where the compatibility of a given $\tau_0 \in \mathcal{C}(n)$ is determined through the following steps.

1. Set $j = 0$.

2. If $j = n + 1$, return that τ_0 is incompatible. Otherwise, continue to the next step.

3. Construct a count vector $\boldsymbol{v}$ as follows (recall (2.2)). Let v_{10} be a free parameter, and set $v_{11} = j - v_{10}$. Then (solving (4.1)) we set

$$v_{01} = v_{10} - n\tau_0, \qquad v_{00} = n - j - v_{10} + n\tau_0. \tag{4.3}$$

4. Find the smallest v_{10} that leads to a possible table $\boldsymbol{v}$ given the observed data $\boldsymbol{n}$. We elaborate upon how this can be done Lemma 4.5, below. If no table is possible, increment j by 1 and return to (2).

5. Permutation test the count vector $\boldsymbol{v}$ coming from the choice of v_{10} in the previous step using Algorithm 2.2 with the given value of α. If $\boldsymbol{v}$ is accepted, return that τ_0 is compatible. If $\boldsymbol{v}$ is rejected and $v_{10} \geq 1$ or $v_{01} \geq 1$, increase j by 1 and return to (2). If $\boldsymbol{v}$ is rejected and $v_{10} = v_{01} = 0$, additionally permutation test the count vector given by $v_{10} = 1$. If it is possible and accepted, terminate the algorithm and declare τ_0 compatible. Otherwise, increment j by 1 and return to (2).

The following lemma determines the possible potential outcome tables, given some observed data, in a certain one-parameter family of tables. We remark that it shows Step (4) of Algorithm 4.4 can be completed in constant time.

Lemma 4.5. *Fix $j, v_{10} \in \mathbb{Z}_{\geq 0}$ and $\tau_0 \in \mathcal{C}$, and consider the potential outcome vector*

$$\boldsymbol{v} = (j - v_{10}, v_{10}, v_{10} - n\tau_0, n - j - v_{10} + n\tau_0). \tag{4.4}$$

Given observed data $\boldsymbol{n}$, a necessary condition for $\boldsymbol{v}$ to be possible is that all of the following inequalities hold:

$$j \geq n\tau_0 + n_{01}, \quad j \geq n_{11}, \quad n \geq j + n_{10}, \quad n_{11} + n\tau_0 + n_{10} + n_{01} \geq j. \tag{4.5}$$

In this case, the possible values of v_{10} are given by the interval

$$\max(0, n\tau_0, j - n_{11} - n_{01}, n_{11} + n_{01} + n\tau_0 - j) \leq v_{10} \leq \min(j, n_{11} + n_{00}, n_{10} + n_{01} + n\tau_0, n + n\tau_0 - j), \tag{4.6}$$

which may be empty.

Our simulations using Algorithm 6.2 show significant improvements over the results reported in [RH15, LD16], as measured by the number of permutation tests required to compute $\mathcal{I}(\boldsymbol{n})$ for particular examples of $\boldsymbol{n}$. We will report on these simulations in a future version of this article.

5 Missing Data

We now consider a more general situation in which the realized outcomes Y_i may not be observed for some subjects, with no restriction on the distribution of the unobserved Y_i. More precisely, we suppose that instead of observing $(\boldsymbol{Y}, \boldsymbol{Z})$, the experimenter observes $(\boldsymbol{M}, \boldsymbol{Z})$, where

$$\boldsymbol{M} \in \{-1, 0, 1\}^n, \qquad M_i = -J_i + (1 - J_i)Y_i, \tag{5.1}$$

and $\boldsymbol{J} \in \{1, 0\}^n$. Here $\boldsymbol{J}$ is a random variable which may have an arbitrary distribution, and in particular may depend on $\boldsymbol{Y}$ and $\boldsymbol{Z}$. The indices i such that $J_i = 0$ denote realized outcomes that are observed, while those such that $J_i = 1$ indicate missing data. Hence, we have $M_i = -1$ if the realized outcome for the i-th subject is not observed.

We now construct exact confidence intervals for $\tau(\boldsymbol{y})$ given $(\boldsymbol{M}, \boldsymbol{Z})$ by considering, roughly speaking, the two extremal imputations of the missing data (leading to the smallest and largest estimates of $\tau(\boldsymbol{y})$).

Definition 5.1. Given a observed data $(\boldsymbol{M}, \boldsymbol{Z})$, we define the vector $\boldsymbol{Y}^{(+)} \in \{0,1\}^n$ by

$$Y_i^{(+)} = 1 \text{ if } J_i = 1 \text{ and } Z_i = 1, \tag{5.2}$$

$$Y_i^{(+)} = 0 \text{ if } J_i = 1 \text{ and } Z_i = 0, \tag{5.3}$$

$$Y_i^{(+)} = M_i \text{ if } J_i = 0. \tag{5.4}$$

Similarly, we define $\boldsymbol{Y}^{(-)} \in \{0,1\}^n$ by

$$Y_i^{(-)} = 0 \text{ if } J_i = 1 \text{ and } Z_i = 1, \tag{5.5}$$

$$Y_i^{(-)} = 1 \text{ if } J_i = 1 \text{ and } Z_i = 0, \tag{5.6}$$

$$Y_i^{(-)} = M_i \text{ if } J_i = 0. \tag{5.7}$$

Finally, for $\alpha \in (0,1)$, we set

$$\mathcal{I}_\alpha^\circ(\boldsymbol{M}, \boldsymbol{Z}) = \big[L_\alpha(\boldsymbol{Y}^{(-)}, \boldsymbol{Z}), U_\alpha(\boldsymbol{Y}^{(+)}, \boldsymbol{Z})\big]. \tag{5.8}$$

The following proposition is proved in Section 9.

Proposition 5.2. *Fix potential outcomes $\boldsymbol{y}$ and $\alpha \in (0,1)$. Then*

$$\mathbb{P}\big(\tau(\boldsymbol{y}) \in \mathcal{I}_\alpha^\circ(\boldsymbol{M}, \boldsymbol{Z})\big) \geq 1 - \alpha, \tag{5.9}$$

where the probability is with respect to the variable $\boldsymbol{Z}$.

Remark 5.3. The intervals $\mathcal{I}_\alpha^\circ$ can be used to lift our hypothesis that n is even, in the following way. Given a group of $2m-1$ subjects, insert a fictitious subject to create a group of $2m$ subjects, randomize to equal groups, and construct $\mathcal{I}_\alpha^\circ$ by treating the data from the fictitious subject as missing. An argument similar to the one that proves Proposition 5.2 shows that this covers the sample average treatment effect for the original group of $2m-1$ subjects with probability at least $1-\alpha$. Such intervals will slightly sacrifice precision relative to the intervals $\mathcal{I}_\alpha$ applied to an unbalanced design with groups of m and $m-1$. (This loss of precision is straightforwardly quantified and decays with rate n^{-1}.) However, the only provably correct construction of $\mathcal{I}_\alpha$ in the unbalanced setting requires $O(n^3)$ permutation tests (as discussed in Section 1), while $\mathcal{I}_\alpha^\circ$ can be constructed in $O(n \log n)$ permutation tests by Theorem 4.1. Hence, there is a trade-off between precision and computational practicality, which may favor using the $\mathcal{I}_\alpha^\circ$ construction when n is large.

6 Monte Carlo Intervals

In Section 4, we constructed the interval $\mathcal{I}_\alpha$ using exact permutation tests (Algorithm 2.2). In this section, we show how to construct confidence intervals using approximate Monte Carlo permutation tests, which are far less computationally demanding.

We begin by defining an approximate permutation test.

Algorithm 6.1. Take as input $\alpha \in (0,1)$, $\varepsilon \in (0,\alpha)$, $K \in \mathbb{Z}_{>0}$, a potential outcome table $\boldsymbol{w}$, and a vector $\boldsymbol{n} \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$. The algorithm will return a binary decision, either accept or reject.

Let $\boldsymbol{Z}^{(1)}, \dots, \boldsymbol{Z}^{(K)}$ be a sequence of independent, identically distributed random vectors, with common distribution equal to that of $\boldsymbol{Z}$. Define random variables $V_1, \dots, V_K$ by

$$V_i = \mathbb{1}\Big(\big|T(\boldsymbol{w}, \boldsymbol{Z}^{(i)}) - \tau(\boldsymbol{w})\big| \geq \big|T(\boldsymbol{n}) - \tau(\boldsymbol{w})\big|\Big), \qquad i \in [\![1, K]\!], \tag{6.1}$$

where $T(\boldsymbol{w}, \boldsymbol{Z}^{(i)})$ is to defined to be $T(\widetilde{\boldsymbol{Y}}, \boldsymbol{Z}^{(i)})$ and $\widetilde{\boldsymbol{Y}}$ is the observed data arising from $\boldsymbol{w}$ and $\boldsymbol{Z}$. Set

$$S = \frac{1}{K}\sum_{i=1}^{K} V_i. \tag{6.2}$$

Simulate a realization of S by simulating realizations of the K variables $\boldsymbol{Z}^{(1)}, \dots, \boldsymbol{Z}^{(K)}$, and constructing the V_i and S accordingly. Return the decision accept if $S + \varepsilon \geq \alpha$, and return the decision reject otherwise.

Algorithm 6.2. This algorithm takes as input $\alpha \in (0,1)$, $\varepsilon \in (0,\alpha)$, and a vector $\boldsymbol{n} \in \mathbb{Z}_{\geq 0}^{\{0,1\}^2}$. It returns an interval $\mathcal{I}_{\alpha,\varepsilon,K}(\boldsymbol{n}) = \big[L_{\alpha,\varepsilon,K}(\boldsymbol{n}), U_{\alpha,\varepsilon,K}(\boldsymbol{n})\big]$. The algorithm is identical to Section 4, except every use of an exact permutation test (Algorithm 2.2) is replaced with an approximate permutation test with the given parameters (Algorithm 6.1).

Given observed data $(\boldsymbol{Y}, \boldsymbol{Z})$ we define the interval $\mathcal{I}_{\alpha,\varepsilon,K}(\boldsymbol{Y}, \boldsymbol{Z})$ to be the interval $\mathcal{I}_{\alpha,\varepsilon,K}(\boldsymbol{n})$ returned by Algorithm 6.2 for the observed count vector $\boldsymbol{n}$ corresponding to $(\boldsymbol{Y}, \boldsymbol{Z})$.

The first part of the next proposition shows that the intervals generated by Algorithm 6.2 are guaranteed to cover $\tau(\boldsymbol{y})$ with probability $1-\alpha$, if the parameters of the algorithm are chosen correctly. The second part shows that these intervals can be made to approximate the deterministic intervals $\mathcal{I}_\alpha$ arbitrarily well.

Proposition 6.3. *Fix a potential outcome table $\boldsymbol{y}$ and $\alpha \in (0,1)$.*

1. *If $K \geq \varepsilon^{-2}\log(4/\varepsilon)$, then*

$$\mathbb{P}\big(\tau(\boldsymbol{y}) \in \mathcal{I}_{\alpha-\varepsilon,\varepsilon,K}(\boldsymbol{Y}, \boldsymbol{Z})\big) \geq 1 - \alpha, \tag{6.3}$$

where the probability is with respect to the variable $\boldsymbol{Z}$.

2. *We have*

$$\mathbb{P}\big(\mathcal{I}_{\alpha-\varepsilon,\varepsilon,K}(\boldsymbol{Y}, \boldsymbol{Z}) \subset \mathcal{I}_{\alpha-3\varepsilon}\big) \geq 1 - 8(n+1)\lfloor \log_2(n+1) + 2\rfloor \exp(-K\varepsilon^2). \tag{6.4}$$

Proposition 6.3 clarifies two important points. First, the number of Monte Carlo samples K of $T(\boldsymbol{w}, \widetilde{Y})$ should increase roughly quadratically in ε^{-1} in order to guarantee that the Monte Carlo confidence intervals cover at the appropriate rate. Second, the number of samples should increase logarithmically in n to guarantee that the Monte Carlo intervals are not unduly large.

We close with two notes on the practical implementation of Algorithm 6.2. First, it is trivially parallelizable, for example by dividing the computation of the V_i in (6.1) across distinct processors. Second, for clarity we have stated Algorithm 6.2 using a fixed number of K samples for each potential outcome table $\boldsymbol{w}$ to be tested using Algorithm 6.1. A simple modification is to stop sampling variables V_i and reject $\boldsymbol{w}$ when the probability that $S + \varepsilon > \alpha$ becomes sufficiently small in Algorithm 6.1. In general, this will require fewer samples than the lower bound for K given in Proposition 6.3, which is essentially a worst-case bound for tables $\boldsymbol{w}$ such that $p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) \approx \alpha$. Such an adaptive algorithm is straightforward to construct by inspecting the concentration inequalities used in the proof of Proposition 6.3.

7 Proofs for Section 3

The following definition will be used in the proofs of our results. For $k \in \mathbb{Z}_{>0}$, we denote

$$k^{-1}\mathbb{Z} = \left\{ \frac{j}{k} : j \in \mathbb{Z} \right\}, \quad k^{-1}(2\mathbb{Z}) = \left\{ \frac{2j}{k} : j \in \mathbb{Z} \right\}, \quad k^{-1}(2\mathbb{Z}+1) = \left\{ \frac{2j+1}{k} : j \in \mathbb{Z} \right\}. \tag{7.1}$$

Definition 7.1. Let $f \colon \mathbb{Z} \to \mathbb{R}_{\geq 0}$ be a probability mass function. We say that f is *symmetric about* k_0 for some $k_0 \in 2^{-1}\mathbb{Z}$ if

$$f(k_0 - x) = f(k_0 + x) \text{ for all } x \in 2^{-1}\mathbb{Z} \text{ such that } k_0 + x \in \mathbb{Z}. \tag{7.2}$$

If f is symmetric about k_0, we say that f is *decreasing away from* k_0 if

$$f(k_0 + y) \leq f(k_0 + x) \text{ for all } x, y \in 2^{-1}\mathbb{Z} \text{ such that } y > x \text{ and } k_0 + x, k_0 + y \in \mathbb{Z}. \tag{7.3}$$

We now suppose that $\sum_{k=-\infty}^{\infty} k f(k)$ converges absolutely, so that the mean of the distribution corresponding to f exists. In this case, we say that f is *symmetric–decreasing* (abbreviated SD) if f is symmetric about $k_0 = \sum_{k=-\infty}^{\infty} k f(k)$, and decreasing away from k_0.

Given $k \in \mathbb{Z}_{>0}$ and $k_0 \in 2^{-1}\mathbb{Z}$, we say that a probability mass function $f \colon k^{-1}\mathbb{Z} \to \mathbb{R}_{\geq 0}$ is symmetric about $k^{-1}k_0 \in (2k)^{-1}\mathbb{Z}$ if the function $g \colon \mathbb{Z} \to \mathbb{R}_{\geq 0}$ defined by $g(x) = f(kx)$ is symmetric about k_0. Similarly, we say that f is SD if $g(x) = f(kx)$ is.

7.1 Proof of Lemma 3.1

We begin with two preliminary lemmas.

Lemma 7.2. *For any potential outcome table $\boldsymbol{w}$, the pmf of $T(\boldsymbol{w}, \boldsymbol{Z})$ is symmetric about $\mathbb{E}\big[T(\boldsymbol{w}, \boldsymbol{Z})\big]$.*

Proof. Let $\boldsymbol{Z}^\dagger$ be the random vector defined by $Z_i^\dagger = 1 - Z_i$. Then for any randomization given by $\boldsymbol{Z}$, the vector $\boldsymbol{Z}^\dagger$ represents a randomization that exchanges the control and treatment groups. We have

$$\frac{1}{2}\Big(T(\boldsymbol{w}, \boldsymbol{Z}) + T(\boldsymbol{w}, \boldsymbol{Z}^\dagger)\Big) = \frac{1}{2m}\sum_{i=1}^{n}\big(w_j(1) - w_j(2)\big) = \mathbb{E}\big[T(\boldsymbol{w}, \boldsymbol{Z})\big]. \tag{7.4}$$

Let f be the pmf of $T(\boldsymbol{w}, \boldsymbol{Z})$, which is supported on $m^{-1}\mathbb{Z}$. Since $\boldsymbol{Z}$ and $\boldsymbol{Z}^\dagger$ have the same distribution, we deduce from (7.4) that

$$f\Big(\mathbb{E}\big[T(\boldsymbol{w}, \boldsymbol{Z})\big] - j\Big) = f\Big(\mathbb{E}\big[T(\boldsymbol{w}, \boldsymbol{Z})\big] + j\Big) \tag{7.5}$$

for all $j \in n^{-1}\mathbb{Z}$ such that $\mathbb{E}\big[T(\boldsymbol{w}, \boldsymbol{Z})\big] + j \in m^{-1}\mathbb{Z}$, which completes the proof. □

Lemma 7.3. *Let $n = 2m$ for some $m \in \mathbb{Z}_{>0}$, and fix observed data $\boldsymbol{Y}, \boldsymbol{Z}$. Then then there exists a potential outcome table $\boldsymbol{w}$ such that $\tau(\boldsymbol{w}) = T(\boldsymbol{Y}, \boldsymbol{Z})$ and $\boldsymbol{w}$ is possible given $\boldsymbol{Y}, \boldsymbol{Z}$.*

Proof. We have $T(\boldsymbol{Y},\boldsymbol{Z}) = m^{-1}(n_{11} - n_{01})$. Let $\boldsymbol{w}$ be a potential outcome table with count vector $\boldsymbol{v} = (0, n_{11} + n_{00}, n_{10} + n_{01}, 0)$. Then using

$$n_{11} + n_{10} = n_{01} + n_{00} = m, \tag{7.6}$$

we compute

$$\tau(\boldsymbol{w}) = \frac{1}{n}(n_{11} + n_{00} - n_{10} - n_{01}) \tag{7.7}$$

$$= \frac{1}{n}\big(n_{11} + (m - n_{01}) - (m - n_{11}) - n_{01}\big) = \frac{2n_{11} - 2n_{01}}{n} = T(\boldsymbol{Y},\boldsymbol{Z}). \tag{7.8}$$

This completes the proof. □

Proof of Lemma 3.1. Let $\boldsymbol{w}$ be the potential outcome table provided by Lemma 7.3, so that $\tau(\boldsymbol{w}) = T(\boldsymbol{Y},\boldsymbol{Z})$ and $\mathcal{I}_\alpha(\boldsymbol{n}) = \mathcal{I}_\alpha(\boldsymbol{Y},\boldsymbol{Z})$. Then the distribution of $T(\boldsymbol{w},\boldsymbol{Z})$ is symmetric about $T(\boldsymbol{Y},\boldsymbol{Z})$, so $T(\boldsymbol{Y},\boldsymbol{Z})$ is the median of the distribution. Then by (2.12), we have $p(\boldsymbol{w},\boldsymbol{Y},\boldsymbol{Z}) = 1$, which shows that $\tau(\boldsymbol{w})$ will always be accepted in the second step of Algorithm 2.3 when searching for L_α and U_α, if $\tau(\boldsymbol{w})$ is ever considered. This completes the proof. □

7.2 Proof of Proposition 3.2

Proof of Proposition 3.2. The event $\big\{\tau(\boldsymbol{y}) \notin \mathcal{I}_\alpha(\boldsymbol{y},\boldsymbol{Z})\big)\big\}$ is contained in the event where Algorithm 2.3 rejects $\tau(\boldsymbol{y})$, which is in turn contained in the event where

$$\mathbb{P}\left(\big|T(\widetilde{\boldsymbol{Y}},\widetilde{\boldsymbol{Z}}) - \tau(\boldsymbol{y})\big| \ge \big|T(\boldsymbol{Y},\boldsymbol{Z}) - \tau(\boldsymbol{y})\big|\right) < \alpha. \tag{7.9}$$

Here probability is with respect to $\widetilde{\boldsymbol{Z}}$, and we consider the left side of (7.9) as a function of $T(\boldsymbol{Y},\boldsymbol{Z})$. Let x denote the smallest real number such that

$$\mathbb{P}\left(\big|T(\widetilde{\boldsymbol{Y}},\widetilde{\boldsymbol{Z}}) - \tau(\boldsymbol{y})\big| \ge x\right) < \alpha, \tag{7.10}$$

which exists because $T(\widetilde{\boldsymbol{Y}},\widetilde{\boldsymbol{Z}})$ is a discrete random variable. Then the event where (7.9) occurs is the same as the event where $\big|T(\boldsymbol{Y},\boldsymbol{Z}) - \tau(\boldsymbol{y})\big| \ge x$. By the definition of x, and the fact that $(\boldsymbol{Y},\boldsymbol{Z})$ and $(\widetilde{\boldsymbol{Y}},\widetilde{\boldsymbol{Z}})$ are identically distributed, this event has probability at most α. This completes the proof. □

7.3 Proof of Proposition 3.3

Proof of Proposition 3.3. Let $\mathcal{P}$ be the set of all sequences of pairs $\big((q_i,r_i)\big)_{i=1}^m$ such that $q_i, r_i \in [\![1,n]\!]$ for all $i \in [\![1,m]\!]$ and

$$\{q_1, q_2, \dots, q_m, r_1, r_2, \dots r_m\} = \{1, 2, \dots, n\}. \tag{7.11}$$

Observe that every element of $[\![1,n]\!]$ appears exactly once in such a sequence as some q_i or r_i.

We now perform complete randomization on the group of n subjects in the following way. First, let $P = \big((Q_i,R_i)\big)_{i=1}^m$ be a random variable uniformly distributed on the set $\mathcal{P}$. Hence, for all $i \in [\![1,m]\!]$, Q_i and R_i are random variables taking values in $[\![1,n]\!]$. We define a random vector

$\widetilde{\boldsymbol{Z}} \in \{0,1\}^n$ as follows. For each $i \in [\![1, N]\!]$, set

$$(\widetilde{Z}_{Q_i}, \widetilde{Z}_{R_i}) = (1,0) \text{ if } B_i = 1, \qquad (\widetilde{Z}_{Q_i}, \widetilde{Z}_{R_i}) = (0,1) \text{ if } B_i = 0. \tag{7.12}$$

One straightforwardly checks that the resulting $\widetilde{\boldsymbol{Z}}$ has a uniform distribution on randomizations of the n subjects to two equal groups. That is, $\widetilde{\boldsymbol{Z}}$ has the same distribution as $\boldsymbol{Z}$.

Next, let $\boldsymbol{w}$ by any potential outcome table. We will condition on P and derive a concentration inequality for the resulting conditional distribution of $T(\boldsymbol{w}, \widetilde{\boldsymbol{Z}})$. First note that the conditional mean of $T(\boldsymbol{w}, \widetilde{\boldsymbol{Z}})$ is

$$\mathbb{E}\big[T(\boldsymbol{w}, \widetilde{\boldsymbol{Z}})\big|P\big] = \tau(\boldsymbol{w}).$$

Next, observe that the variables

$$A_i = Z_{Q_i} y_{Q_i}(1) - (1 - Z_{Q_i}) y_{Q_i}(2) + Z_{R_i} y_{R_i}(1) - (1 - Z_{R_i}) y_{R_i}(2) \tag{7.13}$$

defined for $i \in [\![1, m]\!]$ are conditionally independent, after conditioning on P, and the conditional distribution of $T(\boldsymbol{w}, \widetilde{\boldsymbol{Z}})$ is given by

$$\frac{1}{m} \sum_{i=1}^{m} A_i. \tag{7.14}$$

Observe that we have $A_i \in \{-1, 0, 1\}$.

Now recall that Hoeffding's inequality states that given independent random variables $\{X_i\}_{i=1}^n$ such that $a \le X \le b$, we have

$$\mathbb{P}\Big(\big|S_n - \mathbb{E}[S_n]\big| \ge t\Big) \le 2 \exp\left(-\frac{2t^2}{n(b-a)^2}\right) \tag{7.15}$$

for all $t > 0$, where

$$S_n = X_1 + \cdots + X_n. \tag{7.16}$$

We apply Hoeffding's inequality to the sequence $\{m^{-1} A_i\}_{i=1}^m$ with $a = -m^{-1}$ and $b = m^{-1}$, after conditioning on P, yielding

$$\mathbb{P}\Big(\big|T(\boldsymbol{w}, \widetilde{\boldsymbol{Z}}) - \tau(\boldsymbol{w})\big| \ge t \Big| P\Big) \le 2 \exp\left(-\frac{t^2 n}{8}\right). \tag{7.17}$$

Because this inequality is true conditional on every realization of $\mathcal{P}$, we have

$$\mathbb{P}\Big(\big|T(\boldsymbol{w}, \widetilde{\boldsymbol{Z}}) - \tau(\boldsymbol{w})\big| \ge t\Big) \le 2 \exp\left(-\frac{t^2 n}{8}\right). \tag{7.18}$$

Inserting $t = \big|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau(\boldsymbol{w})\big|$, we get

$$\mathbb{P}\Big(\big|T(\boldsymbol{w}, \widetilde{\boldsymbol{Z}}) - \tau(\boldsymbol{w})\big| \ge \big|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau(\boldsymbol{w})\big|\Big) \le 2 \exp\left(-\frac{\big|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau(\boldsymbol{w})\big|^2 n}{8}\right). \tag{7.19}$$

To reject $\boldsymbol{w}$, we must have

$$\mathbb{P}\Big(\big|T(\boldsymbol{w}, \widetilde{\boldsymbol{Z}}) - \tau(\boldsymbol{w})\big| \ge \big|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau(\boldsymbol{w})\big|\Big) < \alpha. \tag{7.20}$$

We introduce the shorthand

$$\delta = \big|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau(\boldsymbol{w})\big|. \tag{7.21}$$

Hence, we solve for

$$2\exp\left(-\frac{\delta^2 n}{8}\right) < \alpha, \tag{7.22}$$

giving

$$\delta \geq \sqrt{\frac{8\log(2/\alpha)}{n}}. \tag{7.23}$$

Therefore the permutation test will reject all values of τ_0 with distance at least

$$\sqrt{\frac{8\log(2/\alpha)}{n}} \tag{7.24}$$

from $T(\boldsymbol{Y}, \boldsymbol{Z})$. This implies the conclusion. □

8 Proofs for Section 4

8.1 Proof of Theorem 4.1

Lemma 8.1. *Algorithm 4.2 terminates in at most* $\lfloor \log_2(k_2 - k_1 + 1) + 2\rfloor$ *evaluations of* f.

Proof. Without loss of generality, we may suppose that $k_1 = 1$. Further, by setting $f(x) = 1$ for $x > k_2$, we may consider f as a function on $[\![1, 2^d + 1]\!]$, where $d = \lfloor \log_2(k_2) + 1\rfloor$ satisfies $2^d \geq k_2$. Then the number of integers in the interval $[a, b]$ after each evaluation of f follows the sequence $2^{d-1}+1, \dots, 2^1+1, 2$ as the algorithm progresses, with a possible additional evaluation of f if $a = 1$ or $b = 2^d + 1$ at the final step. Then at at worst $d+1$ evaluations are needed in total, which proves the theorem. □

We now prove Theorem 4.1 assuming the other results stated in Section 4, which are proved below.

Proof of Theorem 4.1. We will show that Algorithm 4.4 returns $\mathcal{I}_\alpha$ in the required number of permutation tests. It suffices to show that Algorithm 4.4 returns $U_\alpha^{(1)}(\boldsymbol{n}) = U_\alpha(\boldsymbol{n})$ for the upper bound in $2(n+1)\lfloor \log_2(n+1) + 2\rfloor$ permutation tests, since the reasoning for the lower bound is analogous.

Let $C_+ = \max\big(\mathcal{C}(\boldsymbol{n})\big)$, where we recall that $\mathcal{C}(\boldsymbol{n})$ was defined in (2.10). By Lemma 3.1, we have $U_\alpha(\boldsymbol{n}) \in \big[T(\boldsymbol{n}), C_+\big]$. Recall that by Lemma 2.5, the set of compatible elements

$$\mathcal{J}_\alpha(\boldsymbol{n}) = \{\tau_0 \in \mathcal{C}(\boldsymbol{n}) : \text{there exists a table } \boldsymbol{u} \text{ such that } \tau(\boldsymbol{u}) = \tau_0 \text{ and } p(\boldsymbol{n}, \boldsymbol{u}) \geq \alpha\} \tag{8.1}$$

is an interval equal to $I_\alpha(\boldsymbol{n})$, so that

$$\big[T(\boldsymbol{n}), C_+\big] \cap J_\alpha(\boldsymbol{n}) = \big[T(\boldsymbol{n}), U_\alpha(\boldsymbol{n})\big]. \tag{8.2}$$

Then it suffices to show that the f given in Algorithm 4.4 satisfies $f(x) = 0$ if $x/n \in \mathcal{J}_\alpha$ and $f(x) = 1$ otherwise, and that $f(x)$ can be computed in at most $2(n+1)$ permutation tests for every x. In this case, the binary search indicated in Algorithm 4.4 will return $U_\alpha(\boldsymbol{n})$ in a total $2(n+1)\lfloor \log_2(n+1) + 2\rfloor$ permutation tests, as desired.

We begin with the claim that $f(x) = 0$ if $x/n \in \mathcal{J}_\alpha$ and $f(x) = 1$ otherwise. If a vector $\boldsymbol{v}$ considered in Step (5) is compatible (that is, satisfies $p(\boldsymbol{n}, \boldsymbol{v}) \geq \alpha$), then the associated τ_0 is compatible (lies in $\mathcal{J}_\alpha(\boldsymbol{n})$), so f correctly declares τ_0 compatible. Otherwise, if $j = n+1$ and Step (2) declares τ_0 incompatible, we must show that every possible potential outcome count vector $\boldsymbol{u}$ with $\tau(\boldsymbol{u}) = \tau_0$ is incompatible.

Fix an arbitrary $\boldsymbol{u}$. We must show that $p(\boldsymbol{n}, \boldsymbol{u}) < \alpha$. If $\boldsymbol{u}$ was tested for compatibility in Step (5), then this was already established. Otherwise, let $j = u_{11} + u_{01}$, and let v_{01} be the value determined in Step (4). Let $\boldsymbol{v}$ be the vector determined by v_{01} through the equalities in (3). If τ_0 was declared incompatible by the algorithm, then $\boldsymbol{v}$ was declared incompatible. Hence $p(\boldsymbol{n}, \boldsymbol{v}) < \alpha$. Using this inequality and Lemma 4.3, we deduce that

$$p(\boldsymbol{n}, \boldsymbol{u}) \leq p(\boldsymbol{n}, \boldsymbol{v}) < \alpha, \tag{8.3}$$

since $\boldsymbol{v} = \boldsymbol{u} + k(1, -1, -1, 1)$ for some $k \in \mathbb{Z}_{>0}$. Hence, $\boldsymbol{u}$ is incompatible if $\boldsymbol{v}$ is. We have shown that evaluating f provides a valid method for checking the compatibility of τ_0.

Finally, the claim that f can be evaluated using $2(n+1)$ permutation tests follows from the fact that examining a given $j \in [\![0, n]\!]$ requires at most two permutation tests, as described in Step (5). This completes the proof. $\square$

8.2 Proof of Lemma 4.3

We first provide some intuition for the proof of Lemma 4.3. We will see that the sampling distribution for the Neyman estimator $T(\boldsymbol{w})$ of a generic table $\boldsymbol{w}$ is SD on $m^{-1}\mathbb{Z}$. If all tables $\boldsymbol{w}$ had this property, the proof of Lemma 4.3 would be fairly straightforward (by applying Lemma 8.2 below). However, there are exceptional tables that do not, due to parity issues. Whenever we have a table of the form $\boldsymbol{v} = (a, 0, 0, b)$ with $a + b = n = 2m$, the estimator $T(\boldsymbol{w}, \boldsymbol{Z})$ is supported on $m^{-1}(2\mathbb{Z})$ or $m^{-1}(2\mathbb{Z}+1)$ instead of $m^{-1}\mathbb{Z}$.[4] Most of our effort goes into showing that this is the only such bad case.

Lemma 8.2. *Fix observed data* $\boldsymbol{n}$*, and a count vector*

$$\boldsymbol{v} = (v_{11}, v_{10}, v_{01}, v_{00}). \tag{8.4}$$

Suppose $\min(v_{10}, v_{01}) \geq 1$*, and define*

$$\boldsymbol{v}' = (v_{11} + 1, v_{10} - 1, v_{01} - 1, v_{00} + 1). \tag{8.5}$$

Set

$$\boldsymbol{v}^\circ = (v_{11}, v_{10} - 1, v_{01} - 1, v_{00}), \tag{8.6}$$

and suppose that the pmf of $T(\boldsymbol{v}^\circ, \boldsymbol{Z})$ *is SD on the lattice* $(m-1)^{-1}\mathbb{Z}$*, where here we let* $\boldsymbol{Z}$ *be uniformly distributed over* $\mathcal{Z}(n-2)$*. Then*

$$p(\boldsymbol{n}, \boldsymbol{v}') \geq p(\boldsymbol{n}, \boldsymbol{v}). \tag{8.7}$$

Proof. We will couple the distributions of $T(\boldsymbol{v}, \boldsymbol{Z})$ and $T(\boldsymbol{v}', \boldsymbol{Z})$ in the following way. Let $\boldsymbol{w}$ be a potential outcome table with count vector $\boldsymbol{v}$, and label the subjects so that $\boldsymbol{w}_1 = (1, 0)$ and $\boldsymbol{w}_2 = (0, 1)$. Let $\boldsymbol{w}'$ be a potential outcome table such that $\boldsymbol{w}'_1 = (1, 1)$ and $\boldsymbol{w}'_2 = (0, 0)$, and

[4] Here $T(\boldsymbol{w}, Z)$ is defined using the observed values from a randomized experiment with underlying potential outcome table $\boldsymbol{w}$, where the observed values are defined as in (2.5).

$\boldsymbol{w}_i' = \boldsymbol{w}_i$ for $i \geq 3$. Then $\boldsymbol{w}'$ has the count vector $\boldsymbol{v}'$. Let $\widetilde{\boldsymbol{Y}}$ be the observed outcome vector for $\boldsymbol{w}$, let $\widetilde{\boldsymbol{Y}}'$ be the observed outcome vector for $\boldsymbol{w}'$, and recall from (2.12) that

$$p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) = \mathbb{P}\left(\left|T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) - \tau_0\right| \geq \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right|\right). \tag{8.8}$$

by definition.

We may suppose $T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0 \neq 0$, otherwise $p(\boldsymbol{n}, \boldsymbol{v}') = p(\boldsymbol{n}, \boldsymbol{v}) = 1$ and the claim (8.7) is trivially true. Then (8.8) implies

$$p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) = \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) - \tau_0 \geq \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right|\right) \tag{8.9}$$

$$+ \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) - \tau_0 \leq -\left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right|\right). \tag{8.10}$$

To complete the proof, it suffices to show that each of the probabilities (8.9) and (8.10) increases as $\boldsymbol{w}$ is changed to $\boldsymbol{w}'$. Observe that

$$\mathbb{E}\big[T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}})\big] = \tau(\boldsymbol{w}) = \tau_0, \tag{8.11}$$

and that $T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}})$ is symmetric about its mean, by Lemma 7.2. Then by symmetry, it suffices to show that (8.9) is increases when $\boldsymbol{w}$ goes to $\boldsymbol{w}'$.

As a preliminary observation, note that $\tau_0 \in \mathcal{C}$ takes on values in the lattice $n^{-1}\mathbb{Z}$, while T takes on values in $m^{-1}\mathbb{Z} = 2n^{-1}\mathbb{Z}$. However, we are considering the probability

$$\mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) \geq \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0\right), \tag{8.12}$$

from (8.9), and one checks that $\left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0$ is an element of $m^{-1}\mathbb{Z}$ regardless of the value of τ_0, since $T \in m^{-1}\mathbb{Z}$.

Now, given $T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}})$ for some fixed realization of $\widetilde{\boldsymbol{Z}}$, we consider what happens to its value as $\boldsymbol{w}$ changes to $\boldsymbol{w}'$, so that $\widetilde{\boldsymbol{Y}}$ is replaced by $\widetilde{\boldsymbol{Y}}'$. There are three cases.

1. $\widetilde{Z}_1 = \widetilde{Z}_2$. Then the first two subjects are in the same group. The change $(1,0) \mapsto (1,1)$ and $(0,1) \mapsto (0,0)$ leaves T invariant in this case.

2. $\widetilde{Z}_1 = 1$ and $\widetilde{Z}_2 = 0$. The change of $(1,0) \mapsto (1,1)$ for the first subject does not affect T. The change $(0,1) \mapsto (0,0)$ for the second subject increases T by m^{-1}.

3. $\widetilde{Z}_1 = 0$ and $\widetilde{Z}_2 = 1$. The change of $(1,0) \mapsto (1,1)$ for the first subject decreases T by m^{-1}. The change $(0,1) \mapsto (0,0)$ for the second subject does not affect T.

In the conclusion, the value of T changes by 0 or $\pm m^{-1}$. Then for $a \in \mathbb{Z}$, we have

$$\mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}', \widetilde{\boldsymbol{Z}}) = \frac{a}{m}\right) = \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \frac{a-1}{m} \text{ and } (\widetilde{Z}_1, \widetilde{Z}_2) = (1,0)\right) \tag{8.13}$$

$$+ \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \frac{a}{m} \text{ and } \widetilde{Z}_1 = \widetilde{Z}_2\right) \tag{8.14}$$

$$+ \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \frac{a+1}{m} \text{ and } (\widetilde{Z}_1, \widetilde{Z}_2) = (0,1)\right). \tag{8.15}$$

As noted below (8.10), it suffices to show that

$$\mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}', \widetilde{\boldsymbol{Z}}) \geq \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0\right) \geq \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) \geq \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0\right), \tag{8.16}$$

since this implies that (8.9) is increasing if $\boldsymbol{v}$ is replaced by $\boldsymbol{v}'$. Using (8.13), we get

$$\mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}', \widetilde{\boldsymbol{Z}}) \geq \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0\right) \tag{8.17}$$

$$= \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) \geq \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0 + m^{-1}\right) \tag{8.18}$$

$$+ \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0 \text{ and } \widetilde{Z}_1 = \widetilde{Z}_2\right) \tag{8.19}$$

$$+ \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0 \text{ and } (\widetilde{Z}_1, \widetilde{Z}_2) = (1,0)\right) \tag{8.20}$$

$$+ \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0 - m^{-1} \text{ and } (\widetilde{Z}_1, \widetilde{Z}_2) = (1,0)\right). \tag{8.21}$$

Using the previous equality, we see that (8.16) is equivalent to the statement that

$$\mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0 \text{ and } \widetilde{Z}_1 = \widetilde{Z}_2\right) \tag{8.22}$$

$$+ \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0 \text{ and } (\widetilde{Z}_1, \widetilde{Z}_2) = (1,0)\right) \tag{8.23}$$

$$+ \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0 - m^{-1} \text{ and } (\widetilde{Z}_1, \widetilde{Z}_2) = (1,0)\right) \tag{8.24}$$

$$\geq \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0\right), \tag{8.25}$$

which rearranges to

$$\mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| - m^{-1} + \tau_0 \text{ and } (\widetilde{Z}_1, \widetilde{Z}_2) = (1,0)\right) \tag{8.26}$$

$$\geq \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) = \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0 \text{ and } (\widetilde{Z}_1, \widetilde{Z}_2) = (0,1)\right). \tag{8.27}$$

Let $\widetilde{\boldsymbol{Z}}^\circ = (\widetilde{Z}_3, \dots \widetilde{Z}_n)$, let $\boldsymbol{w}^\circ = (\boldsymbol{w}_3, \dots, \boldsymbol{w}_n)$, and let $\widetilde{\boldsymbol{Y}}^\circ$ be the corresponding vector of observed outcomes. Observe that $\boldsymbol{v}^\circ$ is the count vector for $\boldsymbol{w}^\circ$. Then (8.26) may be rewritten as

$$\mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}^\circ, \widetilde{\boldsymbol{Z}}^\circ) = A - (m-1)^{-1}\right) \geq \mathbb{P}\left(T(\widetilde{\boldsymbol{Y}}^\circ, \widetilde{\boldsymbol{Z}}^\circ) = A\right). \tag{8.28}$$

where

$$A = \frac{m}{m-1} \cdot \left(\left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right| + \tau_0\right) \in (m-1)^{-1}\mathbb{Z}, \tag{8.29}$$

and we used the fact that the net contribution to $T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}})$ from the first two subjects is zero, since we assumed that $\boldsymbol{w}_1 = (1,0)$ and $\boldsymbol{w}_2 = (0,1)$. Observe that

$$A > \frac{m}{m-1} \cdot \tau_0 = \mathbb{E}\left[T\big(\widetilde{\boldsymbol{Y}}^\circ, \widetilde{\boldsymbol{Z}}^\circ\big)\right], \tag{8.30}$$

since we assumed earlier that $T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0 \neq 0$. By hypothesis, the pmf of $T\big(\widetilde{\boldsymbol{Y}}^\circ, \widetilde{\boldsymbol{Z}}^\circ\big)$ is SD, since it corresponds to the count vector $\boldsymbol{v}^\circ$. This fact, and the fact that A is strictly greater than the mean by (8.30), together imply that (8.28) holds. Since (8.28) is equivalent to (8.16), we see that (8.16) holds, which completes the proof. □

We next prove some lemmas to set up an induction.

Lemma 8.3. *Let X_1 be SD on $m^{-1}\mathbb{Z}$, and let X_2 be uniformly distributed on $\{0, m^{-1}\}$. Then $X_1 + X_2$ is SD on $m^{-1}\mathbb{Z}$.*

Proof. It suffices to show that $m(X_1 + X_2)$ is SD on $\mathbb{Z}$. Denote the pdf of mX_1 by f_1. Then the pdf g of $m(X_1 + X_2)$ is given by $g(x) = \frac{1}{2}\big(f_1(x) + f_1(x-1)\big)$. Suppose that mX_1 is symmetric about $a \in \mathbb{Z}$. Then it is clear that $g(x)$ is symmetric about $a + \frac{1}{2}$. To show that $g(x)$ is decreasing, it suffices to show that for any $b \geq a + 1/2$, we have $g(b) \geq g(b+1)$. We find

$$g(b) = \frac{1}{2}\big(f_1(b) + f_1(b-1)\big) \geq \frac{1}{2}\big(f_1(b+1) + f_1(b)\big) = g(b+1), \tag{8.31}$$

as desired. In the previous equation, we used $f(b) \geq f(b+1)$, which is true since f is SD, and $f_1(b-1) \geq f(b)$. The latter is true by the SD property of f_1 if $b \geq a+1$. If $b = a + 1/2$, then this follows from the symmetric of f about a. □

Lemma 8.4. *Fix a potential outcome count vector*

$$\boldsymbol{v} = (v_{11}, v_{10}, v_{01}, v_{00}), \tag{8.32}$$

and suppose at least one of the following conditions holds.

1. *We have $v_{10} \geq 2$, and the pmf for $T(\boldsymbol{v} - \boldsymbol{a}, \boldsymbol{Z})$ is SD on $(m-1)^{-1}\mathbb{Z}$ for every choice of*

$$\boldsymbol{a} \in \big\{(1,1,0,0), (0,2,0,0), (0,1,1,0), (0,1,0,1)\big\} \tag{8.33}$$

   such that $\boldsymbol{v} - \boldsymbol{a}$ has nonnegative entries.

2. *We have $v_{01} \geq 2$, and the pmf for $T(\boldsymbol{v} - \boldsymbol{a}, \boldsymbol{Z})$ is SD on $(m-1)^{-1}\mathbb{Z}$ for every choice of*

$$\boldsymbol{a} \in \big\{(1,0,1,0), (0,1,1,0), (0,0,2,0), (0,0,1,1)\big\} \tag{8.34}$$

   such that $\boldsymbol{v} - \boldsymbol{a}$ has nonnegative entries.

Then the pmf of $T(\boldsymbol{v}, \boldsymbol{Z})$ is SD on $m^{-1}\mathbb{Z}$.

Proof. We prove the claim only for the case $v_{10} \geq 2$, since the proof in the case that $v_{01} \geq 2$ is similar. By relabeling, we may suppose that the first subject has the potential outcome vector $\boldsymbol{w}_1 = (1, 0)$. We sample $\boldsymbol{Z}$ in the following way. First, choose subject j from $\{2, \dots, n\}$ uniformly at random and consider the pair (w_1, w_j). Then sample (Z_1, Z_j) uniformly at random from the assignments $(Z_1 = 1, Z_j = 0)$ and $(Z_1 = 0, Z_j = 1)$. Finally, assign the remaining $n-2$ subjects to groups by an independent randomization to equal groups (chosen uniformly at random). We will show that conditional distribution of $T(\boldsymbol{v}, \boldsymbol{Z})$, conditional on the choice of j, is SD with a mean of $\mathbb{E}\big[T(\boldsymbol{v}, \boldsymbol{Z})\big]$. This suffices to prove the theorem, since the unconditional pmf for $T(\boldsymbol{v}, \boldsymbol{Z})$ is the weighted sum of such conditional pmfs. There are four cases.

1. $\boldsymbol{w}_j = (1, 1)$. Then $w_1(1) - w_j(2) = 0$, and $w_j(1) - w_1(2) = 1$, so the distribution of

$$Z_1 w_1(1) + (1 - Z_1) w_1(2) + Z_j w_j(1) + (1 - Z_j) w_j(2) \tag{8.35}$$

   is uniform on the set $\{0, 1\}$. Further, the contribution to T from the other $n-2$ subjects, after the independent randomization is also symmetric decreasing on $m^{-1}\mathbb{Z}$, by assumption. Then the sum of these (conditionally) independent variables is SD with mean is $\mathbb{E}\big[T(\boldsymbol{v}, \boldsymbol{Z})\big]$, by Lemma 8.3.

2. $\boldsymbol{w}_j = (1, 0)$. One checks that the distribution (8.35) is constant and equal to 1, and that the same reasoning as in the previous case applies, since the sum of a random variable that is SD and the constant 1 is still SD.

3. $\boldsymbol{w}_j = (0, 1)$. The distribution of (8.35) is constant and equal to 0, and the same reasoning as in the previous point applies.

4. $\boldsymbol{w}_j = (0, 0)$. The distribution of (8.35) is uniform on $\{0, 1\}$, and we can apply Lemma 8.3.

This completes the proof. □

Lemma 8.5. *Fix any potential outcome count vector of the form*

$$\boldsymbol{v} = (v_{11}, 0, 0, v_{00}). \tag{8.36}$$

Then the pmf of $T(\boldsymbol{v}, \boldsymbol{Z})$ is SD on $m^{-1}(2\mathbb{Z})$ if v_{11} is even. It is SD on $m^{-1}(2\mathbb{Z}+1)$ if v_{11} is odd.

Remark 8.6. Note that this statement concerns the lattices $m^{-1}(2\mathbb{Z})$ and $m^{-1}(2\mathbb{Z}+1)$, instead of the lattice $m^{-1}\mathbb{Z}$, due to the parity issue mentioned in the introduction to this subsection.

Proof. We prove the claim for all $n \in \mathbb{Z}_{\geq 0}$ by induction. The base case $n = 0$ is trivial. For the induction step, we suppose that the claim holds for $n - 2$ and will show it is true for n. Fix an arbitrary $\boldsymbol{v}$ of the form (8.36) with $v_{11} + v_{00} = n$. We suppose that $v_{11} \geq 2$. When $v_{11} = 0$, the conclusion is trivial, and when $v_{11} = 1$, we can apply the reasoning below with v_{00} instead of v_{11}.

By relabeling, we may suppose $\boldsymbol{w}_1 = (1, 1)$. We sample $\boldsymbol{Z}$ as in the previous proof by choosing a partner $\boldsymbol{w}_j$ for $\boldsymbol{w}_1$ uniformly at random, randomizing the pair $\boldsymbol{w}_1$ and $\boldsymbol{w}_j$ to treatment and control, and independently randomizing the other $n - 2$ subjects to equal groups.

We consider first the case that $\boldsymbol{w}_2 = (1, 1)$. Then the distribution of

$$Z_1 w_1(1) + (1 - Z_1) w_1(2) + Z_j w_j(1) + (1 - Z_j) w_j(2) \tag{8.37}$$

conditional on the choice of partner and that fact that subjects 1 and 2 are assigned to different groups is constant and equal to 0. This implies we are done by induction, since $v_{11} - 2$, the number of $(1, 1)$ potential outcome vectors in the remaining $n - 2$ subjects, has the same parity as v_{11}, and the sampling distributions of

$$\sum_{i=1}^{n} Z_i w_i(1) + (1 - Z_i) w_i(2) \tag{8.38}$$

and

$$\sum_{i=3}^{n} Z_i w_i(1) + (1 - Z_i) w_i(2) \tag{8.39}$$

are then both supported and SD on either $m^{-1}(2\mathbb{Z})$ or $m^{-1}(2\mathbb{Z}+1)$, as desired.

We next consider the case where $\boldsymbol{w}_2 = (0, 0)$. Then the conditional distribution of (8.37) is uniform on $\{-1, 1\}$. We are again done by induction, since there are $v_{11} - 1$ potential outcome vectors of the form $(1, 1)$ in the remaining $n - 2$ subjects, and this number has the opposite parity to v_{11}. Reasoning as in the proof of Lemma 8.3, we see that the sums (8.38) and (8.39) are such that one is supported and SD on $m^{-1}(2\mathbb{Z})$ and one is supported and SD on $m^{-1}(2\mathbb{Z}+1)$, as desired.

Combining the conclusions of these two cases completes the proof. □

Lemma 8.7. *Let $n = 2m$ for some $m \in \mathbb{Z}_{>0}$. Fix a potential outcome count vector*

$$\boldsymbol{v} = (v_{11}, 1, 0, v_{00}) \text{ or } (v_{11}, 0, 1, v_{00}). \tag{8.40}$$

Then the pmf of $T(\boldsymbol{v}, \boldsymbol{Z})$ is SD on $m^{-1}\mathbb{Z}$.

Proof. By relabeling, we may suppose $\boldsymbol{w}_1 = (1, 1)$, and we resample $\boldsymbol{Z}$ as in the proof of Lemma 8.4 by first choosing uniformly at random a partner $\boldsymbol{w}_j$ for $\boldsymbol{w}_1$. By Lemma 8.5, the distribution of $(m-1)T$ for the remaining $n-2$ people is SD on $2\mathbb{Z}$ or $2\mathbb{Z}+1$, where T is evaluated just with respect to this group fo $n-2$ subjects. Similarly, the distribution of $2T$ for the partnership is always uniform on $\{0, 1\}$. Then the pmf for mT for the entire set of subjects is SD on $\mathbb{Z}$ by Lemma 8.3. □

The proof of the following lemma is somewhat computational, so we defer it to the next subsection.

Lemma 8.8. *Let $n = 2m$ for some $m \in \mathbb{Z}_{>0}$. Fix a potential outcome count vector*

$$\boldsymbol{v} = (v_{11}, 2, 0, v_{00}). \tag{8.41}$$

Then the pmf of $T(\boldsymbol{v}, \boldsymbol{Z})$ is SD on $m^{-1}\mathbb{Z}$.

Corollary 8.9. *Let $n = 2m$ for some $m \in \mathbb{Z}_{>0}$. Fix a potential outcome count vector*

$$\boldsymbol{v} = (v_{11}, 1, 1, v_{00}) \text{ or } (v_{11}, 0, 2, v_{00}). \tag{8.42}$$

Then the pmf of $T(\boldsymbol{v}, \boldsymbol{Z})$ is SD on $m^{-1}\mathbb{Z}$.

Proof. The pmfs of $T(\boldsymbol{v}, \boldsymbol{Z})$ for the given vectors are translates of the pmf corresponding to $T(\boldsymbol{v}, \boldsymbol{Z})$ for

$$\boldsymbol{v} = (v_{11}, 2, 0, v_{00}), \tag{8.43}$$

so we are done by the previous lemma. □

Lemma 8.10. *Let $n = 2m$ for some $m \in \mathbb{Z}_{>0}$. Fix any potential outcome count vector*

$$\boldsymbol{v} = (v_{11}, v_{10}, v_{01}, v_{00}), \tag{8.44}$$

and suppose $v_{10} + v_{01} \geq 1$. Then the pmf of $T(\boldsymbol{v}, \boldsymbol{Z})$ is SD on $m^{-1}\mathbb{Z}$.

Proof. We induct on n. The base case $n = 0$ is trivial. For the induction step, suppose that the claim is true for $n-2$, and fix any $\boldsymbol{v}$ corresponding to a potential outcome table with n subjects such that $\min(v_{10}, v_{01}) \geq 1$. If $(v_{10}, v_{01}) = (1, 0)$ or $(v_{10}, v_{01}) = (0, 1)$, we are done by Lemma 8.7. If $v_{01} + v_{01} = 2$, we are done by Lemma 8.8 and Corollary 8.9. In the case $v_{01} + v_{01} > 2$, we use Lemma 8.4 to show the claim is true for the given $\boldsymbol{v}$.

In this case, suppose first that $v_{10} \geq 2$ and $v_{01} \geq 1$. Then by the induction hypothesis, Lemma 8.7, Lemma 8.8, and Corollary 8.9, the first condition given in the statement of Lemma 8.4 holds. The case where $v_{10} \leq 1$ but $v_{01} \geq 2$ is analogous. This completes the proof. □

Proof of Lemma 4.3. This is an immediate consequence of combining Lemma 8.10 and Lemma 8.2. □

8.3 Proof of Lemma 8.8

We first recall a formula of Copas [Cop73]. Given a potential outcome table $\boldsymbol{w}$ and a randomization $\boldsymbol{Z}$, we define a *treatment count vector* $\boldsymbol{x}$ by

$$\boldsymbol{x} = (x_{11}, x_{10}, x_{01}, x_{00}), \qquad x_{ab} = \sum_{j=1}^{n} Z_j \mathbb{1}\{\boldsymbol{w}_j = (a,b)\}. \tag{8.45}$$

Note that

$$x_{11} + x_{10} + x_{01} + x_{00} = m, \tag{8.46}$$

if $n = 2m$ and we randomize into equal groups.

Lemma 8.11 ([Cop73])**.** *Fix a potential outcome count vector $\boldsymbol{v}$. For any $s_0, s_1 \in \mathbb{Z}_{\geq 0}$, the probability of observing a treatment count vector $\boldsymbol{x}$ such that*

$$s_1 = x_{11} + x_{10}, \qquad s_0 = (v_{11} - x_{11}) + (v_{01} - x_{01}) \tag{8.47}$$

is

$$p(s_1, s_0) = C_{\boldsymbol{v}} \sum_{x=-\infty}^{\infty} \binom{v_{11}}{x} \binom{v_{10}}{s_1 - x} \binom{v_{01}}{v_{11} + v_{01} - s_0 - x} \binom{v_{00}}{m - v_{11} - s_1 - v_{01} + s_0 + x}, \tag{8.48}$$

where $C_{\boldsymbol{v}} > 0$ is a constant depending only on $\boldsymbol{v}$.

Proof of Lemma 8.8. We explicitly compute the distribution of $T(\boldsymbol{v}, \boldsymbol{Z})$. We first note that direct computation shows that $m\mathbb{E}\big[T(\boldsymbol{v}, \boldsymbol{Z})\big] = 1$, and we know that the distribution of $T(\boldsymbol{v}, \boldsymbol{Z})$ is symmetric about its mean by Lemma 7.2.

Adopt the notation of Lemma 8.11. Then

$$s_1 = x_{11} + x_{10}, \qquad s_0 = v_{11} - x_{11}. \tag{8.49}$$

With p defined as in (8.48), and using the assumed form of $\boldsymbol{v}$, we have

$$p(s_1, s_0) = C_{\boldsymbol{v}} \sum_{x} \binom{v_{11}}{x} \binom{2}{s_1 - x} \binom{0}{v_{11} - s_0 - x} \binom{v_{00}}{m - v_{11} - s_1 - v_{01} + s_0 + x} \tag{8.50}$$

$$= C_{\boldsymbol{v}} \sum_{x} \binom{v_{11}}{x} \binom{2}{x_{11} + x_{10} - x} \binom{0}{x_{11} - x} \binom{v_{00}}{m - v_{11} - s_1 + s_0 + x}. \tag{8.51}$$

For this to be nonzero, we must have $x = x_{11}$. Using this and (8.49), we get

$$p(s_1, s_0) = p(x_{11}, x_{10}) = C_{\boldsymbol{v}} \binom{v_{11}}{x_{11}} \binom{2}{x_{10}} \binom{v_{00}}{m - x_{11} - x_{10}}. \tag{8.52}$$

We abbreviate $j = mT(\boldsymbol{v}, \boldsymbol{Z}) - 1$ (centering by subtracting the expectation). Then (using $v_{01} = 0$)

$$j = s_1 - s_0 - 1 = x_{11} + x_{10} - v_{11} + x_{11} - 1 = 2x_{11} + x_{10} - v_{11} - 1. \tag{8.53}$$

This yields

$$x_{11} = \frac{j + v_{11} - x_{10} + 1}{2}. \tag{8.54}$$

We also have

$$\frac{v_{00}+v_{11}}{2}+1=m. \tag{8.55}$$

Using (8.54) and (8.55) in (8.52), we get

$$p(x_{11},x_{10})=C_{\boldsymbol{v}}\binom{2}{x_{10}}\binom{v_{11}}{\frac{v_{11}}{2}+\frac{1-x_{10}+j}{2}}\binom{v_{00}}{\frac{v_{00}}{2}+\frac{1-x_{10}-j}{2}}. \tag{8.56}$$

We now consider two cases, depending on the parity of v_{11}.

Case I: v_{11} is even. Then v_{00} is also even, since n is even. There are two subcases, depending on the parity of j. Suppose that j is even. Then parity considerations from (8.54) force $x_{10}=1$. Then the probability mass function becomes

$$p(j)=2C_{\boldsymbol{v}}\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j}{2}}\binom{v_{00}}{\frac{v_{00}}{2}-\frac{j}{2}}. \tag{8.57}$$

When j is odd, both $x_{10}=0$ and $x_{10}=2$ are possible, and the pmf is

$$p(j)=C_{\boldsymbol{v}}\binom{v_{11}}{\frac{v_{11}}{2}+\frac{1+j}{2}}\binom{v_{00}}{\frac{v_{00}}{2}+\frac{1-j}{2}}+C_{\boldsymbol{v}}\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j-1}{2}}\binom{v_{00}}{\frac{v_{00}}{2}+\frac{-1-j}{2}}. \tag{8.58}$$

Now it is a straightforward computation to show that $p(j)$ is decreasing for $j\geq 0$. We will show that $p(j)\geq p(j+1)$ for all $j\geq 0$. First, suppose that j is even. Then, dividing through by $C_{\boldsymbol{v}}$. we must show that

$$2\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j}{2}}\binom{v_{00}}{\frac{v_{00}}{2}-\frac{j}{2}}\geq\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j+2}{2}}\binom{v_{00}}{\frac{v_{00}}{2}-\frac{j}{2}}+\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j}{2}}\binom{v_{00}}{\frac{v_{00}}{2}+\frac{-2-j}{2}}. \tag{8.59}$$

This is clear, since $\binom{n}{k}$ is symmetric and decreasing away from $k=n/2$.

Next, we suppose that j is odd. We want to show

$$\binom{v_{11}}{\frac{v_{11}}{2}+\frac{1+j}{2}}\binom{v_{00}}{\frac{v_{00}}{2}+\frac{1-j}{2}}+\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j-1}{2}}\binom{v_{00}}{\frac{v_{00}}{2}+\frac{-1-j}{2}}\geq 2\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j+1}{2}}\binom{v_{00}}{\frac{v_{00}}{2}-\frac{j+1}{2}}. \tag{8.60}$$

This is again clear, for the same reason.

Case II: v_{11} is odd. Then v_{11} is odd, since n is even. There are two subcases, depending on the parity of j. Suppose that j is odd. Then parity considerations from (8.54) force $x_{10}=1$. Then the probability mass function becomes

$$p(j)=2C_{\boldsymbol{v}}\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j}{2}}\binom{v_{00}}{\frac{v_{00}}{2}-\frac{j}{2}}. \tag{8.61}$$

When j is even, both $x_{10}=0$ and $x_{10}=2$ are possible, and the pmf is

$$p(j)=C_{\boldsymbol{v}}\binom{v_{11}}{\frac{v_{11}}{2}+\frac{1+j}{2}}\binom{v_{00}}{\frac{v_{00}}{2}+\frac{1-j}{2}}+C_{\boldsymbol{v}}\binom{v_{11}}{\frac{v_{11}}{2}+\frac{j-1}{2}}\binom{v_{00}}{\frac{v_{00}}{2}+\frac{-1-j}{2}}. \tag{8.62}$$

The same verification as in the previous case shows that $p(j)$ is decreasing for $j\geq 0$. $\square$

8.4 Proof of Lemma 4.5

Proof of Lemma 4.5. We apply the criterion of Lemma 2.4. The maximum in that statement becomes

$$\max(0, n_{11} - v_{10}, j - v_{10} - n_{01}, j - n\tau_0 - n_{10} - n_{01}). \tag{8.63}$$

Similarly, the minimum becomes

$$\min(j - v_{10}, n_{11}, j - n\tau_0 - n_{01}, n - v_{10} - n_{10} - n_{01}). \tag{8.64}$$

Each argument of the maximum function must be less than each argument of the minimum function. We test the arguments of the maximum from left to right. Starting with 0, we must have

$$j \geq v_{10}, \quad n_{11} \geq 0, \quad j \geq n\tau_0 + n_{01}, \quad n - n_{01} - n_{10} \geq v_{10}, \tag{8.65}$$

where we note that the condition $n_{11} \geq 0$ is always true. Considering $n_{11} - v_{10}$, we get

$$j \geq n_{11}, \quad v_{10} \geq 0, \quad v_{10} \geq n_{11} + n_{01} + n\tau_0 - j, n \geq n_{11} + n_{10} + n_{01}. \tag{8.66}$$

Note that the last condition is always satisfied. Considering $j - v_{10} - n_{01}$, we get

$$n_{10} \geq 0, \quad v_{10} \geq j - n_{11} - n_{01}, \quad v_{10} \geq n\tau_0, \quad n \geq j + n_{10}, \tag{8.67}$$

and the first condition is always true. Finally, considering $j - n\tau_0 - n_{10} - n_{01}$,

$$n\tau_0 + n_{10} + n_{01} \geq v_{10}, \quad n_{11} + n\tau_0 + n_{10} + n_{01} \geq j, \quad n_{10}, \quad n + n\tau_0 - j \geq v_{10}. \tag{8.68}$$

Collecting the previous four displays completes the proof. □

9 Proof of Proposition 5.2

Proof of Proposition 5.2. First, note that by Proposition 3.2, we have

$$\mathbb{P}\big(\tau(\boldsymbol{y}) \in \mathcal{J}_\alpha(\boldsymbol{M}, \boldsymbol{Z})\big) \geq 1 - \alpha, \tag{9.1}$$

where

$$J_\alpha(\boldsymbol{M}, \boldsymbol{Z}) = \bigcup_{\boldsymbol{Y}' \in \mathcal{Y}(\boldsymbol{M})} \mathcal{I}_\alpha(\boldsymbol{Y}', \boldsymbol{Z}), \tag{9.2}$$

and $\mathcal{Y}(\boldsymbol{M})$ consists of all vectors of realized outcomes $\boldsymbol{Y}'$ that could arise from the partially observed vector $\boldsymbol{M}$. To complete the proof of the theorem it then suffices to show that

$$J_\alpha(\boldsymbol{M}, \boldsymbol{Z}) \subset I^\circ_\alpha(\boldsymbol{M}, \boldsymbol{Z}), \tag{9.3}$$

or equivalently,

$$I^\circ_\alpha(\boldsymbol{M}, \boldsymbol{Z})^c \subset J_\alpha(\boldsymbol{M}, \boldsymbol{Z})^c = \bigcap_{\boldsymbol{Y}' \in \mathcal{Y}(\boldsymbol{M})} \mathcal{I}_\alpha(\boldsymbol{Y}', \boldsymbol{Z})^c. \tag{9.4}$$

Fix some $\tau_0 \in I^\circ_\alpha(\boldsymbol{M}, \boldsymbol{Z})^c$. We may suppose that $\tau_0 > U_\alpha(\boldsymbol{Y}^{(+)}, \boldsymbol{Z})$, since the analogous argument in the complementary case is similar. Then by definition, for all $\boldsymbol{w}$ such that $\tau(\boldsymbol{w}) = \tau_0$, $\boldsymbol{w}$ is incompatible with the data $(\boldsymbol{Y}^{(+)}, \boldsymbol{Z})$. We must show that all such $\boldsymbol{w}$ are incompatible with all choices of $(\boldsymbol{Y}', \boldsymbol{Z})$ with $\boldsymbol{Y}' \in \mathcal{Y}(\boldsymbol{M})$. Fix such a $\boldsymbol{w}$ and a choice of $\boldsymbol{Y}'$.

By the definition of $Y^{(+)}$, and since $T(\boldsymbol{Y}^{(+)}, \boldsymbol{Z})$ and $T(\boldsymbol{Y}', \boldsymbol{Z})$ are coupled through $\boldsymbol{Z}$, we have

$$\tau_0 \geq T(\boldsymbol{Y}^{(+)}, \boldsymbol{Z}) \geq T(\boldsymbol{Y}', \boldsymbol{Z}). \tag{9.5}$$

Then

$$\mathbb{P}\left(\left|T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) - \tau_0\right| \geq \left|T(\boldsymbol{Y}, \boldsymbol{Z}) - \tau_0\right|\right) \leq \mathbb{P}\left(\left|T(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{Z}}) - \tau_0\right| \geq \left|T(\boldsymbol{Y}', \boldsymbol{Z}) - \tau_0\right|\right) < \alpha, \tag{9.6}$$

so $\boldsymbol{w}$ is incompatible with $(\boldsymbol{Y}', \boldsymbol{Z})$, as desired. □

10 Proof of Proposition 6.3

Lemma 10.1. *Fix $\varepsilon > 0$ and $K \in \mathbb{Z}_{>0}$. Then*

$$\mathbb{P}\left(\left|p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) - S\right| > \varepsilon\right) \leq 2\exp\left(-K\varepsilon^2\right). \tag{10.1}$$

Proof. We have $\mathbb{E}[V_i] = p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z})$ and $0 \leq V_i \leq 1$ by definition, so Hoeffding's inequality yields

$$\mathbb{P}\left(\left|K \cdot p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) - \sum_{i=1}^{K} V_i\right| > K\varepsilon\right) \leq 2\exp\left(-\frac{(K\varepsilon)^2}{K}\right) = 2\exp\left(-K\varepsilon^2\right), \tag{10.2}$$

as desired. □

Lemma 10.2. *Fix a potential outcome table $\boldsymbol{y}$ and $\alpha \in (0, 1)$.*

1. *With probability at least $1 - 4\exp(-K\varepsilon^2)$, we have $\mathcal{I}_\alpha(\boldsymbol{Y}, \boldsymbol{Z}) \subset \mathcal{I}_{\alpha,\varepsilon,K}(\boldsymbol{Y}, \boldsymbol{Z})$.*
2. *With probability at least $1 - 8(n+1)\lfloor \log_2(n+1) + 2\rfloor \exp(-K\varepsilon^2)$, we have $\mathcal{I}_{\alpha,\varepsilon,K}(\boldsymbol{Y}, \boldsymbol{Z}) \subset \mathcal{I}_{\alpha-2\varepsilon}(\boldsymbol{Y}, \boldsymbol{Z})$*

Proof. We begin with the first claim. Let $\boldsymbol{w}$ be the potential outcome table that Algorithm 4.4 accepts to find $U_\alpha(\boldsymbol{Y}, \boldsymbol{Z})$. The event that $U_\alpha(\boldsymbol{Y}, \boldsymbol{Z}) < U_{\alpha,\varepsilon,K}$ is contained in the event that Algorithm 6.2 rejects $\boldsymbol{w}$ using Algorithm 6.1. This happens when $S + \varepsilon < \alpha$, while we know that $\boldsymbol{p}(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) \geq \alpha$. Then by Lemma 10.1, Algorithm 6.1 rejects $\boldsymbol{w}$ with probability at most $2\exp(-K\varepsilon^2)$. The claim follows after also accounting for the event that $L_\alpha(\boldsymbol{Y}, \boldsymbol{Z}) > L_{\alpha,\varepsilon,K}$ and using a union bound.

For the second claim, to have $\mathcal{I} \subset \mathcal{I}_{\alpha,\varepsilon,K}$, Algorithm 6.2 must reject every potential outcome table $\boldsymbol{w}$ rejected by Algorithm 4.4. By our analysis in the proof of Theorem 4.1, there will be at most $4(n+1)\lfloor \log_2(n+1)+2\rfloor$ such rejections in Algorithm 4.4. Each rejected table in Algorithm 4.4 has $p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) < \alpha - 2\varepsilon$. For the approximate permutation test applied to such a $\boldsymbol{w}$, we bound the acceptance probability as

$$\mathbb{P}(S + \varepsilon \geq \alpha) \leq \mathbb{P}\left(|p(\boldsymbol{w}, \boldsymbol{Y}, \boldsymbol{Z}) - S| > \varepsilon\right) \leq 2\exp(-K\varepsilon^2) \tag{10.3}$$

using Lemma 10.1. This completes the proof after using a union bound. □

Proof of Proposition 6.3. For the first part, it suffices to choose K large enough so that

$$\mathbb{P}(\mathcal{I}_{\alpha-\varepsilon} \subset \mathcal{I}_{\alpha-\varepsilon,\varepsilon,K}) \leq \varepsilon. \tag{10.4}$$

Then the claim follows from Lemma 10.2. The second claim is immediate from Lemma 10.2. □

References

[BB19] Zach Branson and Marie-Abèle Bind. Randomization-based inference for Bernoulli trial experiments and implications for observational studies. *Statistical Methods in Medical Research*, 28(5):1378–1398, 2019.

[Cop73] J. B. Copas. Randomization models for the matched and unmatched 2×2 tables. *Biometrika*, 60(3):467–476, 1973.

[IR15] Guido Imbens and Donald Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

[LD16] Xinran Li and Peng Ding. Exact confidence intervals for the average causal effect on a binary outcome. *Statistics in Medicine*, 35(6):957–960, 2016.

[RH15] Joseph Rigdon and Michael Hudgens. Randomization inference for treatment effects on a binary outcome. *Statistics in Medicine*, 34(6):924–935, 2015.

[Rob88] James Robins. Confidence intervals for causal parameters. *Statistics in medicine*, 7(7):773–785, 1988.

[SS80] Thomas Santner and Mark Snell. Small-sample confidence intervals for $p_1 - p_2$ and p_1/p_2 in 2×2 contingency tables. *Journal of the American Statistical Association*, 75(370):386–394, 1980.

[Was04] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

[Wik22] Wikipedia contributors. Binary search algorithm — Wikipedia, the free encyclopedia, 2022. [Online; accessed 25-October-2022].