TOPICS IN TWO-SAMPLE TESTING

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF STATISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Nelson C. Ray
2013

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Susan P. Holmes) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Persi W. Diaconis)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Bradley Efron)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Jerome H. Friedman)

Approved for the University Committee on Graduate Studies.

Contents

Chapter 1

Stein's method

In this chapter, we present an introduction to Stein's method of exchangeable pairs, which we use to prove the core theoretical result of this thesis: a rate of convergence bound on the Kolmogorov distance between the randomization distribution of the t-statistic and the standard normal distribution.

Due to similarities between this problem and Hoeffding's combinatorial central limit theorem (HCCLT), we first review Stein's proof of the HCCLT via the method of exchangeable pairs.

1.1 Introduction

Stein's method provides a means of bounding the distance between two probability distributions in a given probability metric. When applied with the normal distribution as the target, this results in central-limit-type theorems. Several flavors of Stein's method (e.g. the method of exchangeable pairs) proceed via auxiliary randomization. In Appendix B, we reproduce Stein's [13] proof of the HCCLT.

It will be instructive to follow the proof of the HCCLT because our proof proceeds in a similar fashion but with the following generalizations: an approximate contraction property, less cancellation of terms due to separate estimation of various denominators, and non-unit variance of the random variable of interest. We also refer to auxiliary results in Appendix A.

1.2 Stein's Theorem

Theorem 1.1 bounds the Kolmogorov distance between the distribution of the random variable W and the standard normal distribution in terms of functions of the difference of the exchangeable pair (W, W'). It is applied to the situation where W is the sum of the random diagonal of a matrix to prove the HCCLT. Chen et al. generalize Theorem 1.1 to allow for situations in which the regression condition does not hold exactly, and we later in addition relax the assumption that the W has unit variance.

Theorem 1.1 (Stein). *If W, W' are mean 0, exchangeable random variables with variance 1 satisfying the exact regression condition*

$$\mathbb{E}[W' - W|W] = -\lambda W$$

for some $\lambda \in (0, 1)$, then

$$\begin{aligned}\sup_{w\in\mathbb{R}} |P(W \leq w) - \Phi(w)| &\leq 2\sqrt{\mathbb{E}\left[1 - \frac{1}{2\lambda}E[(W' - W)^2|W]\right]} + (2\pi)^{-1/4}\sqrt{\frac{1}{\lambda}\mathbb{E}|W' - W|^3} \\ &\leq 2\sqrt{\mathrm{Var}(\mathbb{E}[(W' - W)^2|W])} + (2\pi)^{-1/4}\sqrt{\frac{1}{\lambda}\mathbb{E}|W' - W|^3}.\end{aligned}$$

1.3 Hoeffding's Combinatorial CLT

Stein's proof of the HCCLT relies on an application of Theorem 1.1.

Theorem 1.2 (Hoeffding's Combinatorial Central Limit Theorem). *Let $\{a_{ij}\}_{i,j}$ be an $n \times n$ matrix of real-valued entries that is row- and column-centered and scaled* *such that the sums of the squares of its elements equals $n-1$:*

$$\sum_{j=1}^{n} a_{ij} = 0$$
$$\sum_{i=1}^{n} a_{ij} = 0$$
$$\sum_{i=1,j=1}^{n} a_{ij}^2 = n-1$$

Let Π be a random permutation of $\{1,\dots,n\}$ drawn uniformly at random from the set of all permutations:

$$P(\Pi = \pi) = \frac{1}{n!}.$$

Define

$$W = \sum_{i=1}^{n} a_{i\Pi(i)}$$

to be the sum of a random diagonal. Then

$$|P(W \le w) - \Phi(w)| \le \frac{C}{\sqrt{n}} \left[\sqrt{\sum_{i,j=1}^{n} a_{ij}^4} + \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^3} \right].$$

Given fixed data a_{ij}, the HCCLT provides a bound in terms of a universal constant C, a sample-size-dependent term $\frac{1}{\sqrt{n}}$, and a function of the data $\sqrt{\sum_{i,j=1}^{n} a_{ij}^4} + \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^3}$. Thus, given C, we can calculate an explicit bound on the Kolmogorov distance. Consider a sequence of matrices, $a_{ij}^{(n)}$, of dimension $n \times n$. In order to achieve a $\mathcal{O}(n^{-1/2})$ rate of convergence, we require the function of the data to be bounded. However, this will not typically be the case.

Bolthausen [2] was able to prove the following result:

Theorem 1.3 (Bolthausen)**.** *Under the conditions of Theorem 1.2, there is an absolute constant $K > 0$ such that*

$$|P(W \le w) - \Phi(w)| \le K \frac{\sqrt{\sum_{i,j=1}^{n} |a_{ij}|^3}}{n}.$$

Then, given the sequence of matrices $a_{ij}^{(n)}$, the theorem yields a convergence rate of $\mathcal{O}(n^{-1/2})$ as long as $\sqrt{\sum_{i,j=1}^{n} |a_{ij}|^3}/\sqrt{n}$ remains bounded.

Now, we return to generalizing Theorem 1.1.

1.4 Generalized Stein's Theorems

Here, we treat the situation where the regression condition fails to hold exactly. Chen et al. [3] serves as an excellent reference for results of this type.

Definition 1.4 (Approximate Stein Pair)**.** *Let (W, W') be an exchangeable pair. If the pair satisfies the "approximate linear regression condition"*

$$\mathbb{E}[W - W'|W] = \lambda(W - R), \tag{1.1}$$

where R is a variable of small order and $\lambda \in (0, 1)$, then we call (W, W') an approximate Stein pair.

Here we generalize Lemma 5.1 from [3] to the setting of non-unit variance:

Theorem 1.5. *If W, W' are mean 0 exchangeable random variables with variance $\mathbb{E}W^2$ satisfying*

$$\mathbb{E}[W' - W|W] = -\lambda(W - R)$$

for some $\lambda \in (0, 1)$ and some random variable R, then for any $z \in \mathbb{R}$ and $a > 0$,

$$\mathbb{E}[(W' - W)^2 \mathbf{1}_{\{-a \le W' - W \le 0\}} \mathbf{1}_{\{z - a \le W \le z\}}] \le 3a\lambda(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|)$$

and

$$\mathbb{E}[(W' - W)^2 \mathbf{1}_{\{0 \le W' - W \le a\}} \mathbf{1}_{\{z - a \le W \le z\}}] \le 3a\lambda(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|).$$

The following theorem will allow us to prove a $\mathcal{O}(n^{-1/4})$ rate under mild conditions. Generalization of Theorem 5.5 from [3]:

Theorem 1.6. *If W, W' are mean 0 exchangeable random variables with variance $\mathbb{E}W^2$ satisfying*

$$\mathbb{E}[W' - W|W] = -\lambda(W - R)$$

for some $\lambda \in (0,1)$ and some random variable R, then

$$\sup_{z\in\mathbb{R}} |P(W \leq z) - \Phi(z)| \leq (2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|W' - W|^3}{\lambda}} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])}$$
$$+ |\mathbb{E}W^2 - 1| + \mathbb{E}|WR| + \mathbb{E}|R|$$

The following result will let us achieve a rate of $\mathcal{O}(n^{-1/2})$ subject to an additional constraint on the data. Generalization of part of Theorem 5.3 from [3]:

Theorem 1.7. *If W, W' are mean 0 exchangeable random variables with variance $\mathbb{E}W^2$ satisfying*

$$\mathbb{E}[W' - W|W] = -\lambda(W - R)$$

for some $\lambda \in (0,1)$ and some random variable R and $|W' - W| \leq \delta$, then

$$\sup_{z\in\mathbb{R}} |P(W \leq z) - \Phi(z)| \leq \frac{.41\delta^3}{\lambda} + 3\delta(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|) + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])}$$
$$+ |\mathbb{E}W^2 - 1| + \mathbb{E}|WR| + \mathbb{E}|R|$$

Chapter 2

Kolmogorov Distance Bounds

In this chapter, we prove the core theoretical results of this thesis: rate of convergence bounds on the Kolmogorov distance between the randomization distribution of the t-statistic and the standard normal distribution, using Theorems 1.6 and 1.7 of Chapter 1.

2.1 Motivation

Motivated by concerns regarding normality assumptions in the hypothesis being tested, Fisher [7] proposed a nonparametric randomization test. Also known as a permutation test, Fisher applied this novel test to Charles Darwin's *Zea mays* data and noted that the achieved significance level was very similar to that observed in the parametric test. Indeed, Diaconis and Holmes [5] used efficient Gray code based calculations to show that the randomization distribution looked remarkably normal. For more history on the development of randomization procedures, see Zabell [14] or David [4]. Diaconis and Lehmann [6] in their comment on Zabell's paper further expanded on some properties of these randomization tests.

Ludbrook and Dudley [9] have written about the advantages of permutation tests, especially in biomedical research, and outlined two models of statistical inference: the so-called population model, formally introduced by Newman and Pearson [10], and Fisher's randomization model [7]. Add some more on these two models...

Under the randomization model and using the language of triangular arrays, Lehmann [8] proved a weak convergence result of the randomization distribution of the t-statistic to the standard normal distribution, however, there is no known Berry-Esseen type bound for this rate of convergence.

Introduced by Stein [13], the eponymous technique provides a powerful means with which to handle dependencies among collections of random variables, a common criticism of classical Fourier analytic methods. In addition, one can easily obtain bounds on rates of convergence. Bentkus and Götze [1] first obtained a Berry-Esseen bound for Student's statistic in the independent but non-identically distributed setting with additional work by Shao [12].

We use Stein's method of exchangeable pairs to prove a conservative bound of $\mathcal{O}(n^{-1/4})$ on the rate of convergence of the randomization t-distribution to the standard normal distribution. With an additional condition on the data, we are able to obtain a $\mathcal{O}(n^{-1/2})$ rate.

2.2 Set-up

We observe two samples with equal sample size: $S_1 = \{u_i\}_{i=1}^{N}$ and $S_2 = \{u_i\}_{i=N+1}^{2N}$. Since we consider the t-statistic under different permutations, it will be convenient to re-write the sample values relative to the null permutation π_0: $S_1 = \{u_{\pi_0(i)}\}_{i=1}^{N}$ and $S_2 = \{u_{\pi_0(i)}\}_{i=N+1}^{2N}$, where $\pi_0(i) = i$. Under the randomization distribution, where Π is a uniformly chosen permutation, Student's two-sample t-statistic is given by

$$
\begin{aligned}
T_\Pi(\{u_{\Pi(i)}\}_{i=1}^{N}, \{u_{\Pi(i)}\}_{i=N+1}^{2N}) &= \frac{\bar{u}_{1,\Pi} - \bar{u}_{2,\Pi}}{\sqrt{\frac{\frac{1}{N-1}\sum_{i=1}^{N}(u_{\Pi(i)}-\bar{u}_{1,\Pi})^2}{N} + \frac{\frac{1}{N-1}\sum_{i=N+1}^{2N}(u_{\Pi(i)}-\bar{u}_{2,\Pi})^2}{N}}} \\
&= \frac{1}{\sqrt{\frac{N}{N-1}}} \frac{\sum_{i=1}^{N} u_{\Pi(i)} - \sum_{i=N+1}^{2N} u_{\Pi(i)}}{\sqrt{\sum_{i=1}^{N}(u_{\Pi(i)} - \bar{u}_{1,\Pi})^2 + \sum_{i=N+1}^{2N}(u_{\Pi(i)} - \bar{u}_{2,\Pi})^2}} \\
&= \sqrt{\frac{N-1}{N}} \frac{q_\Pi}{d_\Pi},
\end{aligned}
$$

where

$$
\begin{aligned}
q_\Pi &= \left(\sum_{i=1, i\neq I}^{N} u_{\Pi(i)} + u_{\Pi(I)} - \sum_{i=N+1, i\neq J}^{2N} u_{\Pi(i)} - u_{\Pi(J)} \right) \\
d_\Pi &= \sqrt{\sum_{i=1}^{N} (u_{\Pi(i)} - \bar{u}_{1,\Pi})^2 + \sum_{i=N+1}^{2N} (u_{\Pi(i)} - \bar{u}_{2,\Pi})^2} \\
\bar{u}_{1,\Pi} &= \frac{1}{N} \sum_{i=1}^{N} u_{\Pi(i)} \text{ and } \bar{u}_{2,\Pi} = \frac{1}{N} \sum_{i=N+1}^{2N} u_{\Pi(i)}.
\end{aligned}
$$

In order to perform hypothesis testing, we compute the observed value of $T_{\Pi=\pi_0}$ and compare that with the randomization distribution of T_Π. We shall create an exchangeable pair (T_Π, T'_Π) by considering a uniformly random transposition (I, J). WLOG, take $I \leq J$. We apply this transposition to the group labels. Note that if $I, J \in \{1, \ldots, N\}$ or $I, J \in \{N+1, \ldots, 2N\}$ then $T'_\Pi = T_\Pi$, where T'_Π is the t-statistic under this random transposition. That is, the t-statistic is invariant to within-group transpositions: the only changes occur when $1 \leq I \leq N$ and $N+1 \leq J \leq 2N$. With this in mind, let's redefine our transposition to be uniformly at random over the N^2 cases where $1 \leq I \leq N$ and $N+1 \leq J \leq 2N$. Thus,

$$
\begin{aligned}
T'_\Pi(\{u_{\Pi(i)}\}_{i=1}^{N}, \{u_{\Pi(i)}\}_{i=N+1}^{2N}) &= T_{\Pi\circ(I,J)}(\{u_{\Pi\circ(I,J)(i)}\}_{i=1}^{N}, \{u_{\Pi\circ(I,J)(i)}\}_{i=N+1}^{2N}) \\
&= \sqrt{\frac{N-1}{N}} \frac{q'_\Pi}{d'_\Pi} \\
q'_\Pi &= \left(\sum_{i=1, i\neq I}^{N} u_{\Pi(i)} + u_{\Pi(J)} - \sum_{i=N+1, i\neq J}^{2N} u_{\Pi(i)} - u_{\Pi(I)} \right) \\
&= q_\Pi - 2u_{\Pi(I)} + 2u_{\Pi(J)} \\
d'_\Pi &= \sqrt{\sum_{i=1}^{N} (u_{\Pi(i)} - \bar{u}'_{1,\Pi})^2 + \sum_{i=N+1}^{2N} (u_{\Pi(i)} - \bar{u}'_{2,\Pi})^2}.
\end{aligned}
$$

2.3 Assumptions

Recall that the t-statistic is invariant up to sign under affine transformations, so we can mean-center and scale so that $\sum_{i=1}^{2N} u_i = 0$ and $\sum_{i=1}^{2N} u_i^2 = 2N$. The transformation that achieves this centering and scaling is given by

$$u_i \leftarrow \sqrt{\frac{2N}{\sum_{i=1}^{2N}(u_i - \bar{u})^2}}(u_i - \bar{u}), \tag{2.1}$$

so we just assume that the u_i's have already been transformed. This can be seen as a very mild assumption: only $u_i = c$ for all i cannot be scaled in this way.

We also assume that the pooled sample standard deviation is non-zero for all permutations:

$$d_\Pi = \sqrt{\sum_{i=1}^{N}(u_{\Pi(i)} - \bar{u}_{1,\Pi})^2 + \sum_{i=N+1}^{2N}(u_{\Pi(i)} - \bar{u}_{2,\Pi})^2} > 0 \tag{2.2}$$

This estimate is zero if and only if there exists a grouping that is constant in each group. The condition also implies that the sample mean for any group is strictly less than 1 in absolute value. In fact, this assumption subsumes the former.

The mean-centering assumption implies that $\sum_{i=1}^{N} u_{\Pi(i)} = -\sum_{i=N+1}^{2N} u_{\Pi(i)}$ and hence that $\bar{u}_{1,\Pi} = -\bar{u}_{2,\Pi}$ for all Π.

Here we establish an equality with d_Π that will prove easier to work with:

$$\begin{aligned}
d_\Pi^2 &= \sum_{i=1}^{N}(u_{\Pi(i)} - \bar{u}_{1,\Pi})^2 + \sum_{i=N+1}^{2N}(u_{\Pi(i)} - \bar{u}_{2,\Pi})^2 \\
&= \sum_{i=1}^{2N} u_{\Pi(i)}^2 - N\bar{u}_{1,\Pi}^2 - N\bar{u}_{2,\Pi}^2 \\
&= 2N - N\bar{u}_{2,\Pi}^2 - N\bar{u}_{2,\Pi}^2 \\
&= 2N(1 - \bar{u}_{2,\Pi}^2)
\end{aligned}$$

Since $d_\Pi > 0$, it follows that $|\bar{u}_{2,\Pi}| < 1$. Define

$$B = \max_\Pi |\bar{u}_{2,\Pi}| < 1. \tag{2.3}$$

2.4 Preliminaries

Here we collect useful bounds and other results. We include them here rather than in Appendix A because in Chapter **??** we compare the theoretical bounds with simulated results.

In order to bound various moments of $\bar{u}_{2,\Pi}$ under the permutation distribution, we use a result of Serfling's [11]:

Proposition 2.1. *Consider sampling without replacement from a finite list of values* $u_1, \dots, u_{2N}$*. Let* $u_\Delta = \max_i u_i - \min_i u_i$*. Then for* $p > 0$*,*

$$\begin{aligned}\mathbb{E}[\bar{u}^p_{2,\Pi}] &\leq \frac{\Gamma(p/2+1)}{2^{p/2+1}}\left[\frac{N+1}{2N}u^2_\Delta\right]^{p/2}(2N)^{-p/2}\\ &\leq \frac{\Gamma(p/2+1)}{2^{p/2+1}}\left[\frac{N+1}{4N}u^2_\Delta\right]^{p/2}N^{-p/2}\\ &:= f_{c_1}(p)N^{-p/2}.\end{aligned} \tag{2.4}$$

By Assumption (2.3),

$$(d_\Pi)^{-p} = \frac{1}{(2N(1-\bar{u}^2_{2,\Pi}))^{p/2}} \leq \frac{1}{(2N(1-B^2))^{p/2}} := f_{c_2}(p)N^{-p/2}. \tag{2.5}$$

The transposition (I, J) also affects the denominator of T'_Π, and we need to quantify the difference between the denominators of T_Π and T'_Π. Letting $\bar{u}'^2_{2,\Pi}$ denote the sample mean of the second group after the transposition,

$$\begin{aligned}\bar{u}'^2_{2,\Pi} &= \left(\bar{u}_{2,\Pi} - \frac{1}{N}u_{\Pi(J)} + \frac{1}{N}u_{\Pi(I)}\right)^2\\ &= \bar{u}^2_{2,\Pi} + 2\bar{u}_{2,\Pi}\left(-\frac{1}{N}u_{\Pi(J)} + \frac{1}{N}u_{\Pi(I)}\right) + \frac{1}{N^2}(u_{\Pi(I)} - u_{\Pi(J)})^2\end{aligned}$$

We consider the difference

$$
\begin{aligned}
h_\Pi &= d_\Pi^2 - d_\Pi'^2 \\
&= 2N - 2N\bar{u}_{2,\Pi}^2 - 2N + 2N\bar{u}_{2,\Pi}'^2 \\
&= 4\bar{u}_{2,\Pi}(u_{\Pi(I)} - u_{\Pi(J)}) + \frac{2}{N}(u_{\Pi(I)} - u_{\Pi(J)})^2
\end{aligned}
$$

Therefore, by the c_r-inequality,

$$
\begin{aligned}
\mathbb{E}[h_\Pi^p] &= \mathbb{E}\left|4\bar{u}_{2,\Pi}(u_{\Pi(I)} - u_{\Pi(J)}) + \frac{2}{N}(u_{\Pi(I)} - u_{\Pi(J)})^2\right|^p \\
&\leq 2^{p-1}\left(\mathbb{E}|4\bar{u}_{2,\Pi}(u_{\Pi(I)} - u_{\Pi(J)})|^p + \mathbb{E}\left|\frac{2}{N}(u_{\Pi(I)} - u_{\Pi(J)})^2\right|^p\right) \\
&\leq 2^{p-1}\left[(4u_\Delta)^p\mathbb{E}\left|\bar{u}_{2,\Pi}\right|^p + \left(\frac{2}{N}u_\Delta^2\right)^p\right] \\
&\leq 2^{p-1}(4u_\Delta)^p f_{c_1}(p)N^{-p/2} + 2^{p-1}(2u_\Delta^2)^p N^{-p} \\
&:= f_{c_3}(p)N^{-p/2}.
\end{aligned}
\tag{2.6}
$$

Now we establish a bound on the difference $d_\Pi - d_\Pi'$ via a bound on the remainder of a zeroth order Taylor approximation. Write

$$
d_\Pi' = \sqrt{d_\Pi^2 - h_\Pi} = f(h_\Pi) = f(0) + R_0(h_\Pi) = d_\Pi + R_0(h_\Pi)
$$

By Taylor's theorem, the remainder of the zeroth-order expansion takes the form

$$
R_0(h_\Pi) = \frac{f'(\xi_L)}{1}h_\Pi = \frac{-h_\Pi}{2\sqrt{d_\Pi^2 - \xi_L}}, \quad \text{where } \xi_L \in [0, h_\Pi].
$$

We are approximating d_Π' by a constant and bounding the error via a function of the first derivative. This is a sufficient approximation because the squared difference h_Π is not so big relative to the flattening out of the square root function. Now

$$
|d_\Pi - d_\Pi'| \leq |R_0(h_\Pi)| \leq \frac{|h_\Pi|}{2\sqrt{d_\Pi^2 - \xi_L}} \leq \frac{|h_\Pi|}{2\sqrt{d_\Pi^2 - \max(0, h_\Pi)}}
$$

Recall that $h_\Pi = d_\Pi^2 - d_\Pi'^2$, so

$$d_\Pi^2 - \max(0, d_\Pi^2 - d_\Pi'^2) = \begin{cases} d_\Pi^2 & \text{if } d_\Pi^2 - d_\Pi'^2 \le 0 \\ d_\Pi'^2 & \text{if } d_\Pi^2 - d_\Pi'^2 > 0 \end{cases}$$

Therefore,

$$|d_\Pi - d'_\Pi| \le \frac{|h_\Pi|}{2\min(d_\Pi, d'_\Pi)} \le \max\left(\frac{|h_\Pi|}{2d_\Pi}, \frac{|h_\Pi|}{2d'_\Pi}\right) \le \frac{|h_\Pi|}{2d_\Pi} + \frac{|h_\Pi|}{2d'_\Pi}.$$

The important thing to do is to isolate $|h_\Pi|$, which is small in expectation, but not absolutely. By the c_r-inequality,

$$\begin{aligned} \mathbb{E}|d_\Pi - d'_\Pi|^p &\le 2^{p-1}\left(\mathbb{E}\left|\frac{h_\Pi}{2d_\Pi}\right|^p + \mathbb{E}\left|\frac{h_\Pi}{2d'_\Pi}\right|^p\right) \\ &\le 2^{-1}\left(\sqrt{\mathbb{E}[h_\Pi^{2p}]\mathbb{E}[d_\Pi^{-2p}]} + \sqrt{\mathbb{E}[h_\Pi^{2p}]\mathbb{E}[d_\Pi'^{-2p}]}\right) \\ &\le \sqrt{f_{c_3}(2p)N^{-2p/2} f_{c_2}(2p)N^{-2p/2}} \qquad \text{by (2.6) and (2.5)} \\ &:= f_{c_4}(p)N^{-p}. \end{aligned} \tag{2.7}$$

With

$$q_\Pi = N\bar{u}_{1,\Pi} - N\bar{u}_{2,\Pi} = -2N\bar{u}_{2,\Pi}, \tag{2.8}$$

(2.4), and noting that q_Π and q'_Π are exchangeable,

$$\mathbb{E}[q_\Pi'^p] = \mathbb{E}[q_\Pi^p] = \mathbb{E}[(-2N\bar{u}_{2,\Pi})^p] \le 2^p N^p f_{c_1}(p)N^{-p/2} := f_{c_5}(p). \tag{2.9}$$

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^p\right] &\le \sqrt{\mathbb{E}|q'_\Pi|^{2p}\mathbb{E}|d_\Pi d'_\Pi|^{-2p}} \\
&\le \sqrt{\mathbb{E}|q_\Pi|^{2p}\sqrt{\mathbb{E}|d_\Pi|^{-4p}\mathbb{E}|d'_\Pi|^{-4p}}} \\
&= \sqrt{\mathbb{E}|q_\Pi|^{2p}\mathbb{E}|d_\Pi|^{-4p}} \\
&\le \sqrt{f_{c_5}(2p)N^{2p/2}f_{c_2}(4p)N^{-4p/2}} \qquad \text{from (2.9) and (2.5)} \\
&:= f_{c_6}(p)N^{-p/2}. \qquad (2.10)
\end{aligned}$$

2.5 Proof

We proceed to verify the conditions of Theorems 1.6 and 1.7. T_Π and T'_Π are exchangeable by construction:

$$\begin{aligned}
P(\Pi = \pi, \Pi' = \pi') &= P(\Pi' = \pi' | \Pi = \pi) P(\Pi = \pi) \\
&= \frac{1}{N^2}\mathbb{1}_{\{\pi' = \pi\circ(i,j), 1\le i\le N, N+1\le j\le 2N\}} P(\Pi = \pi') \\
&= \frac{1}{N^2}\mathbb{1}_{\{\pi = \pi'\circ(i,j), 1\le i\le N, N+1\le j\le 2N\}} P(\Pi = \pi') \\
&= P(\Pi' = \pi | \Pi = \pi') P(\Pi = \pi') \\
&= P(\Pi = \pi', \Pi' = \pi)
\end{aligned}$$

Since (Π, Π') are exchangeable, $(T_\Pi, T'_\Pi) = (T(\Pi), T(\Pi'))$ are exchangeable as well. T_Π, and thus T'_Π by exchangeability, have mean zero by symmetry. Let π^* identify the permutation that reverses the order of the indices after applying the original permutation π. That is, $\pi^* = (2N, \dots, 1)\circ\pi$. Since indices 1 to N correspond to the first group and $N+1$ to $2N$ to the second, π^* reverses the groups after π, so

$T_{\pi^*} = -T_{\pi}$.

$$\begin{aligned}
P(T_\Pi = t) &= \sum_{\pi : T_\pi = t} P(\Pi = \pi) \\
&= \sum_{\pi : T_\pi = t} P(\Pi = \pi^*) \quad \text{by exchangeability} \\
&= \sum_{\pi^* : T_{\pi^*} = -t} P(\Pi = \pi^*) \quad \text{since } T_{\pi^*} = -T_\pi \text{ and } \pi \;\; \pi^* \text{ is bijective} \\
&= P(T_\Pi = -t)
\end{aligned}$$

To show the approximate regression condition, the difference of our exchangeable pair is given by

$$\begin{aligned}
T'_\Pi - T_\Pi &= \sqrt{\frac{N-1}{N}} \left(\frac{q'_\Pi}{d'_\Pi} - \frac{q_\Pi}{d_\Pi} \right) \\
&= \sqrt{\frac{N-1}{N}} \frac{1}{d_\Pi} \left(q'_\Pi - q_\Pi + q'_\Pi \frac{(d_\Pi - d'_\Pi)}{d'_\Pi} \right) \\
&= \sqrt{\frac{N-1}{N}} \frac{1}{d_\Pi} \left(2u_{\Pi(J)} - 2u_{\Pi(I)} + q'_\Pi \frac{(d_\Pi - d'_\Pi)}{d'_\Pi} \right). \qquad (2.11)
\end{aligned}$$

Note that

$$\begin{aligned}
\sqrt{\frac{N-1}{N}} \mathbb{E} \left[\frac{1}{d_\Pi} (2u_{\Pi(J)} - 2u_{\Pi(I)}) \middle| \Pi = \pi \right] &= \sqrt{\frac{N-1}{N}} \frac{2}{d_\Pi} \frac{1}{N^2} \sum_{I=1}^{N} \sum_{I=N+1}^{2N} (u_{\Pi(J)} - u_{\Pi(I)}) \\
&= -\frac{2}{N} T_\Pi.
\end{aligned}$$

Therefore,

$$\sqrt{\frac{N-1}{N}} \mathbb{E} \left[\frac{1}{d_\Pi} (2u_{\Pi(J)} - 2u_{\Pi(I)}) \middle| \Pi = \pi \right] = \sqrt{\frac{N-1}{N}} \mathbb{E} \left[\frac{1}{d_\Pi} (2u_{\Pi(J)} - 2u_{\Pi(I)}) \middle| T_\Pi \right]$$

and

$$\lambda = \frac{2}{N}.$$

$$\begin{aligned}\mathbb{E}[T'_\Pi - T_\Pi | T_\Pi] &= -\lambda T_\Pi + \sqrt{\frac{N-1}{N}}\mathbb{E}\left[\frac{q'_\Pi}{d_\Pi}\frac{(d_\Pi - d'_\Pi)}{d'_\Pi}\middle| T_\Pi\right] \\ &= -\lambda\left(T_\Pi - \binom{N}{2}\sqrt{\frac{N-1}{N}}\mathbb{E}\left[\frac{q'_\Pi}{d_\Pi}\frac{(d_\Pi - d'_\Pi)}{d'_\Pi}\middle| T_\Pi\right]\right)\end{aligned}$$

so

$$R_\Pi = \binom{N}{2}\sqrt{\frac{N-1}{N}}\frac{1}{d_\Pi}\mathbb{E}\left[q'_\Pi\frac{(d_\Pi - d'_\Pi)}{d'_\Pi}\middle| T_\Pi\right]. \tag{2.12}$$

For convenience, we restate Theorem 1.6 of Chapter 1, taking our random variables W to be the randomization t-statistic T_Π and W' to be its coupled counterpart T'_Π:

Theorem 1.8. *If T_Π, T'_Π are mean 0 exchangeable random variables with variance $\mathbb{E}T_\Pi^2$ satisfying*

$$\mathbb{E}[T'_\Pi - T_\Pi | T_\Pi] = -\lambda(T_\Pi - R_\Pi)$$

for some $\lambda \in (0,1)$ and some random variable R_Π, then

$$\begin{aligned}\sup_{t\in\mathbb{R}}|P(T_\Pi \le t) - \Phi(t)| \le (2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|T'_\Pi - T_\Pi|^3}{\lambda}} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(T'_\Pi - T_\Pi)^2|T_\Pi])} \\ + |\mathbb{E}T_\Pi^2 - 1| + \mathbb{E}|T_\Pi R_\Pi| + \mathbb{E}|R_\Pi|\end{aligned}$$

With the preliminaries in place, we proceed to provide bounds on each term in Theorem 1.6, the proofs of which we defer to Appendix C.

Proposition 2.2. $(2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|T'_\Pi - T_\Pi|^3}{\lambda}} < (2\pi)^{-1/4}c_9 N^{-1/4}$.

Proposition 2.3. $\frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(T'_\Pi - T_\Pi)^2|T_\Pi])} \le N^{-1}c_3\sqrt{20 + 16\frac{\sum_{i=1}^{2N} u_i^4}{N^2}}$

Proposition 2.4. $|\mathbb{E}T_\Pi^2 - 1| \le c_2 N^{-1}$

Proposition 2.5. $\mathbb{E}|T_\Pi R| \le \frac{1}{2}(f_{c_6}(4)f_{c_4}(4))^{1/4}\sqrt{2+2c_1}N^{-1/2}$.

Proposition 2.6. $\mathbb{E}|R| \le \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}$.

The bound in Proposition 2.2 is suboptimal, as it will only allow us to obtain a rate of $\mathcal{O}(n^{-1/4})$. In Section 2.6, we introduce an additional condition to improve upon this rate.

Collecting the results of Propositions 2.4, 2.3, 2.2, 2.6, and 2.5, we have

$$\begin{aligned}\sup_{t\in\mathbb{R}}|P(T_\Pi \le t) - \Phi(t)| &\le (2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|T'_\Pi - T_\Pi|^3}{\lambda}} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(T'_\Pi - T_\Pi)^2|T_\Pi])} \\ &\quad + |\mathbb{E}T_\Pi^2 - 1| + \mathbb{E}|T_\Pi R_\Pi| + \mathbb{E}|R_\Pi| \\ &\le (2\pi)^{-1/4}c_9 N^{-1/4} + N^{-1}c_3\sqrt{20 + 16\frac{\sum_{i=1}^{2N} u_i^4}{N^2}} + c_2 N^{-1} \\ &\quad + \frac{1}{2}(f_{c_6}(4)f_{c_4}(4))^{1/4}\sqrt{2+2c_1}N^{-1/2} + \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}\end{aligned}$$

Note that since $\|x\|_4 \le \|x\|_2$,

$$\sum_{i=1}^{2N} u_i^4 \le \left(\sum_{i=1}^{2N} u_i^2\right)^{4/2} = (2N)^2 = 4N^2.$$

This result is similar to the HCCLT. Given fixed data, we can obtain an explicit upper bound on the Kolmogorov distance between the randomization distribution of our statistic of interest and the standard normal distribution.

2.6 Better Rate

Here, we use Theorem 1.7 to establish a rate of $\mathcal{O}(n^{-1/2})$ with the condition that $|T_\Pi - T'_\Pi| \le \delta$ is $\mathcal{O}(n^{-1/2})$.

From Proposition 2.4, $\mathbb{E}T_\Pi^2 \le c_2 N^{-1} + 1$, and from Proposition 2.6, $\mathbb{E}|R| \le \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}$. If $\delta < c_{10}N^{-1/2}$ for N sufficiently large, applying Theorem 1.7, we see

$$\begin{aligned}\sup_{t\in\mathbb{R}}|P(T_\Pi \le t) - \Phi(t)| &\le \frac{.41\delta^3}{\lambda} + 3\delta\left(\sqrt{\mathbb{E}T_\Pi^2} + \mathbb{E}|R|\right) + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(T'_\Pi - T_\Pi)^2|T_\Pi])} \\ &\quad + |\mathbb{E}T_\Pi^2 - 1| + \mathbb{E}|T_\Pi R| + \mathbb{E}|R| \\ &\le .205c_{10}N^{-1/2} + 3c_{10}N^{-1/2}\left(c_2N^{-1} + 1 + \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}\right) \\ &\quad + N^{-1}c_3\sqrt{20 + 16\frac{\sum_{i=1}^{2N} u_i^4}{N^2}} + c_2N^{-1} \\ &\quad + \frac{1}{2}(f_{c_6}(4)f_{c_4}(4))^{1/4}\sqrt{2+2c_1}N^{-1/2} + \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}.\end{aligned}$$

Again, this result is conditional on the data. We can consider a sequence of vectors $\{u_i^{(2N)}\}$, where each $u_i^{(j)}$ is drawn from some distribution p. As long as all data-dependent functions of the bound are "well-behaved," we shall have the desired rates of convergence such as in [2].

To determine whether $\delta = |T_\Pi - T'_\Pi|$ is $\mathcal{O}(n^{-1/2})$ for reasonable classes of data $\{u_i\}$, recall that

$$T_\Pi(\{u_{\Pi(i)}\}_{i=1}^N, \{u_{\Pi(i)}\}_{i=N+1}^{2N}) = \frac{\bar{u}_{1,\Pi} - \bar{u}_{2,\Pi}}{\sqrt{\frac{\frac{1}{N-1}\sum_{i=1}^N (u_{\Pi(i)} - \bar{u}_{1,\Pi})^2}{N} + \frac{\frac{1}{N-1}\sum_{i=N+1}^{2N} (u_{\Pi(i)} - \bar{u}_{2,\Pi})^2}{N}}}.$$

We need to set $\delta = \max_{\pi,i,j}|T_\pi - T_{\pi\circ(i,j)}|$ so that the bound is tight. This appears to be a daunting optimization problem. There are $(2N)!$ permutations and N^2 possible transpositions (i,j) for each permutation. Well, because the t-statistic is invariant to permutations within groups, there are $\binom{2N}{N}$ (really, $\binom{2N}{N}/2$ because of symmetry) permutations to consider.

We have to solve the maximization problem jointly over T and T'. We can attempt to first maximize over T and then T'. Note that these sequential approaches do not work for general optimization problems.

If we sort the data in ascending order such that the two groups are $\{u_{(i)}\}_{i=1}^N$ and

$\{u_{(i)}\}_{i=N+1}^{2N}$, then it seems like we will have maximized $|T|$. The absolute difference between the sample means of the two groups is maximized, while the pooled sample standard deviation is minimized (is this true? source?).

The transposition that should then maximize $|T - T'|$ is $(1, 2N)$ since it swaps the most different points, decreasing the difference in sample means and increasing the pooled sample standard deviation.

Let π^* be the permutation that sorts the data in ascending order such that $u_{\pi^*(i)} = u_{(i)}$, where $u_{(i)}$ are the order statistics of $\{u_i\}$. Let $i^* = 1$ and $j^* = 2N$.

Conjecture 2.7. *$\delta = \max_{\pi,i,j} |T_\pi - T_{\pi\circ(i,j)}|$ is maximized at $\pi = \pi^*$, $i = i^*$, and $j = j^*$.*

This conjecture has held true under many simulations. We can show that when $u_i = i$,

$$\lim_{n\to\infty} \delta\sqrt{n} = 16\sqrt{6}.$$

Appendix A

Auxiliary Results

The c_r-inequality and following corollary will provide useful bounds to come.

Theorem A.1 (The c_r-inequality)**.** *Let X and Y be random variables and $r > 0$. Suppose that $\mathbb{E}|X|^r < \infty$ and $\mathbb{E}|Y|^r < \infty$. Then*

$$\mathbb{E}|X+Y|^r < c_r(\mathbb{E}|X|^r + \mathbb{E}|Y|^r), \tag{A.1}$$

where $c_r = 1$ when $r \leq 1$ and $c_r = 2^{r-1}$ when $r \geq 1$.

Corollary A.2. *Suppose that* $\mathrm{Var}(X) < \infty$ *and* $\mathrm{Var}(Y) < \infty$*. Then*

$$\mathrm{Var}(X+Y) < 2(\mathrm{Var}(X) + \mathrm{Var}(Y)). \tag{A.2}$$

Proof. This follows immediately by applying Theorem A.1 to the centered random variables $X' = X - \mathbb{E}X$ and $Y' = Y - \mathbb{E}Y$. □

Lemma A.3. *If (W, W') is an exchangeable pair, then $\mathbb{E}g(W, W') = 0$ for all anti-symmetric measurable functions such that the expected value exists.*

Here is a slight generalization of Lemma 2.7 from [3]:

Lemma A.4. *Let (W, W') be an approximate Stein pair and $\Delta = W - W'$. Then*

$$\mathbb{E}W = \mathbb{E}R \quad \textit{and} \quad \mathbb{E}\Delta^2 = 2\lambda\mathbb{E}W^2 - 2\lambda\mathbb{E}WR \quad \textit{if } \mathbb{E}W^2 < \infty. \tag{A.3}$$

Furthermore, when $\mathbb{E}W^2 < \infty$*, for every absolutely continuous function* f *satisfying* $|f(w)| \leq C(1 + |w|)$*, we have*

$$\mathbb{E}Wf(W) = \frac{1}{2\lambda}\mathbb{E}(W - W')(f(W) - f(W')) + \mathbb{E}f(W)R. \tag{A.4}$$

Proof. From (1.1) we have

$$\mathbb{E}[\mathbb{E}[W - W'|W]] = \mathbb{E}\lambda(W - R) = \lambda\mathbb{E}W - \lambda\mathbb{E}R.$$

We also have

$$\mathbb{E}[\mathbb{E}[W - W'|W]] = \mathbb{E}W - \mathbb{E}[\mathbb{E}[W'|W]] = \mathbb{E}W - \mathbb{E}W' = 0$$

using exchangeability. Equating the two expressions yields

$$\mathbb{E}W = \mathbb{E}R$$

As an intermediate computation,

$$\begin{aligned}
\mathbb{E}W'W &= \mathbb{E}[\mathbb{E}[W'W|W]] \\
&= \mathbb{E}[W\mathbb{E}[W'|W]] \\
&= \mathbb{E}[W((1 - \lambda)W + \lambda R)] \quad \text{from (1.1)} \\
&= (1 - \lambda)\mathbb{E}W^2 + \lambda\mathbb{E}WR.
\end{aligned} \tag{A.5}$$

Then

$$\begin{aligned}
\mathbb{E}\Delta^2 &= \mathbb{E}(W - W')^2 \\
&= \mathbb{E}W^2 + \mathbb{E}W'^2 - 2\mathbb{E}W'W \\
&= 2\mathbb{E}W^2 - 2((1 - \lambda)\mathbb{E}W^2 + \lambda\mathbb{E}WR) \quad \text{from (A.5)} \\
&= 2\lambda\mathbb{E}W^2 - 2\lambda\mathbb{E}WR.
\end{aligned} \tag{A.6}$$

By the linear growth assumption on f, $\mathbb{E}g(W, W')$ exists for the antisymmetric function $g(x, y) = (x - y)(f(y) + f(x))$. By Lemma A.3,

$$\begin{aligned}0 &= \mathbb{E}(W - W')(f(W') + f(W))\\ &= \mathbb{E}(W - W')(f(W') - f(W)) + 2\mathbb{E}f(W)(W - W')\\ &= \mathbb{E}(W - W')(f(W') - f(W)) + 2\mathbb{E}[f(W)\mathbb{E}[(W - W')|W]]\\ &= \mathbb{E}(W - W')(f(W') - f(W)) + 2\mathbb{E}f(W)(\lambda(W - R)).\end{aligned}$$

Rearranging the expression yields

$$\mathbb{E}Wf(W) = \frac{1}{2\lambda}\mathbb{E}(W - W')(f(W) - f(W')) + \mathbb{E}f(W)R. \tag{A.7}$$

$\square$

This is just a small part of Lemma 2.4 from [3]:

Lemma A.5. *For a given function $h : \mathbb{R} \to \mathbb{R}$, let f_h be the solution to the Stein equation. If h is absolutely continuous, then*

$$||f_h|| \leq 2||h'||. \tag{A.8}$$

Lemma 2.2 from [3]:

Lemma A.6. *For fixed $z \in \mathbb{R}$ and $\Phi(z) = P(Z \leq z)$, the unique bounded solution $f_z(w)$ of the equation*

$$f'(w) - wf(w) = \mathbf{1}_{\{w \leq z\}} - \Phi(z) \tag{A.9}$$

is given by

$$f_z(w) = \begin{cases} \sqrt{2\pi}e^{w^2/2}\Phi(w)[1 - \Phi(z)] & \textit{if } w \leq z \\ \sqrt{2\pi}e^{w^2/2}\Phi(z)[1 - \Phi(w)] & \textit{if } w > z. \end{cases} \tag{A.10}$$

Part of Lemma 2.3 from [3]:

Lemma A.7. *Let $z \in \mathbb{R}$ and let f_z as in (A.10) Then*

$$|(w + u)f_z(w + u) - (w + v)f_z(w + v)| \leq (|w| + \sqrt{2\pi}/4)(|u| + |v|).$$

Appendix B

Stein's Method Proofs

B.0.1 Proof of Theorem 1.2

Proof. In order to construct our exchangeable pair, we introduce the ordered pair of random variables (I, J) independent of Π that represents a uniformly at random draw from the set of all non-null transpositions:

$$P(I = i, J = j) = \frac{1}{n(n-1)} \quad i, j \in \{1, \dots, n\}, i \neq j. \tag{B.1}$$

Define the random permutation Π' by

$$\Pi'(i) = \Pi \circ (I, J) = \begin{cases} \Pi(J) & i = I \\ \Pi(I) & i = J \\ \Pi(i) & \text{else.} \end{cases} \tag{B.2}$$

We construct our exchangeable pair by defining

$$W' = \sum_{i=1}^{n} a_{i\Pi'(i)} = W - a_{I\Pi(I)} + a_{I\Pi(J)} - a_{J\Pi(J)} + a_{J\Pi(I)}. \tag{B.3}$$

We now verify the contraction property:

$$
\begin{aligned}
\mathbb{E}[W - W'|\Pi] &= \mathbb{E}[a_{I\Pi(I)} - a_{I\Pi(J)} + a_{J\Pi(J)} - a_{J\Pi(I)}|\Pi] \\
&= \frac{2}{n}\sum_{i=1}^{n} a_{i\Pi(i)} - \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(j)} \\
&= \frac{2}{n}W - \frac{2}{n}\frac{1}{n-1}\left[\sum_{i,j=1}^{n} a_{i\Pi(j)} - \sum_{i}^{n} a_{i\Pi(i)}\right] \\
&= \frac{2}{n}W + \frac{2}{n}\frac{1}{n-1}W - \frac{2}{n}\frac{1}{n-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_{i\Pi(j)}\right] \\
&= \frac{2}{n}W\left(1 + \frac{1}{n-1}\right) - 0 \\
&= \frac{2}{n-1}W
\end{aligned}
$$

This satisfies our contraction property with

$$
\lambda = \frac{2}{n-1}. \tag{B.4}
$$

To bound the variance component, compute

$$
\begin{aligned}
\mathbb{E}[(W-W')^2|\Pi] &= \mathbb{E}[(a_{I\Pi(I)} - a_{I\Pi(J)} + a_{J\Pi(J)} - a_{J\Pi(I)})^2|\Pi] \\
&= \mathbb{E}[a^2_{I\Pi(I)} + a^2_{J\Pi(J)} + a^2_{I\Pi(J)} + a^2_{J\Pi(I)} \\
&\quad - 2a_{I\Pi(I)}a_{I\Pi(J)} - 2a_{J\Pi(J)}a_{J\Pi(I)} - 2a_{I\Pi(I)}a_{J\Pi(I)} - 2a_{J\Pi(J)}a_{I\Pi(J)} \\
&\quad + 2a_{I\Pi(I)}a_{J\Pi(J)} + 2a_{I\Pi(J)}a_{J\Pi(I)}|\Pi] \\
&= \frac{2}{n}\sum_{i=1}^{n} a^2_{i\Pi(i)} + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a^2_{i\Pi(j)} \\
&\quad - \frac{4}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(i)}a_{i\Pi(j)} - \frac{4}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(i)}a_{j\Pi(i)} \\
&\quad + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(i)}a_{j\Pi(j)} + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(j)}a_{j\Pi(i)} \\
&= \frac{2}{n}\sum_{i=1}^{n} a^2_{i\Pi(i)} + \frac{2}{n}\frac{1}{n-1}\left(\sum_{i,j=1}^{n} a^2_{i\Pi(j)} - \sum_{i=1}^{n} a^2_{i\Pi(i)}\right) \\
&\quad - \frac{4}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(a_{i\Pi(i)}\sum_{j=1}^{n}\left(a_{i\Pi(j)} + a_{j\Pi(i)}\right) - 2a^2_{i\Pi(i)}\right) \\
&\quad + \frac{2}{n}\frac{1}{n-1}\left(\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right) \\
&= \frac{2}{n}\left(1 - \frac{1}{n-1}\right)\sum_{i=1}^{n} a^2_{i\Pi(i)} + \frac{2}{n} \\
&\quad + \frac{8}{n}\frac{1}{n-1}\sum_{i=1}^{n} a^2_{i\Pi(i)} \\
&\quad + \frac{2}{n}\frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right) - \frac{4}{n}\frac{1}{n-1}\sum_{i=1}^{n} a^2_{i\Pi(i)} \\
&= \frac{2}{n} + \frac{2(n+2)}{n(n-1)}\sum_{i=1}^{n} a^2_{i\Pi(i)} + \frac{2}{n(n-1)}\sum_{i,j=1,i\neq j}^{n}(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)})
\end{aligned}
\tag{B.5}
$$

From (B.5) and corollary A.2,

$$
\begin{aligned}
\mathbb{E}[(W-W')^2|\Pi] &= \operatorname{Var}\left(\frac{2(n+2)}{n(n-1)}\sum_{i=1}^{n} a_{i\Pi(i)}^2\right. \\
&\quad \left. + \frac{2}{n(n-1)}\sum_{i,j=1,i\neq j}^{n}(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)})\right) \\
&\leq 2\left(\frac{4(n+2)^2}{n^2(n-1)^2}\operatorname{Var}\left(\sum_{i=1}^{n} a_{i\Pi(i)}^2\right) + \right. \\
&\quad \left. \frac{4}{n^2(n-1)^2}\operatorname{Var}\left(\sum_{i,j=1,i\neq j}^{n}(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)})\right)\right) \\
&\leq \frac{32}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} a_{i\Pi(i)}^2\right) + \frac{32}{n^4}\operatorname{Var}\left(\sum_{i,j=1,i\neq j}^{n}(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)})\right)
\end{aligned}
\tag{B.6}
$$

for $n \geq 2$ since $n-1 \geq n/2$ $\quad \frac{1}{(n-1)^2} \leq \frac{4}{n^2}$ for $n \geq 2$.

First, we address the first term in (B.6):

$$
\operatorname{Var}\left(\sum_{i=1}^{n} a_{i\Pi(i)}^2\right) = \sum_{i=1}^{n}\operatorname{Var}(a_{i\Pi(i)}^2) + \sum_{i,j=1,i\neq j}^{n}\operatorname{Cov}(a_{i\Pi(i)}^2, a_{j\Pi(j)}^2),
$$

with

$$
\begin{aligned}
\sum_{i,j=1,i\neq j}^{n}\operatorname{Cov}(a_{i\Pi(i)}^2, a_{j\Pi(j)}^2) &= \sum_{i,j=1,i\neq j}^{n}\left(\frac{1}{n(n-1)}\sum_{k,l=1,k\neq l}^{n} a_{ik}^2 a_{jl}^2 - \left(\frac{1}{n}\sum_{k} a_{ik}^2\right)\left(\frac{1}{n}\sum_{l} a_{jl}^2\right)\right) \\
&= \sum_{i,j=1,i\neq j}^{n}\left(\frac{1}{n(n-1)}\sum_{k,l=1}^{n} a_{ik}^2 a_{jl}^2 - \frac{1}{n^2}\sum_{k}\sum_{l} a_{ik}^2 a_{jl}^2 - \frac{1}{n(n-1)}\sum_{k} a_{ik}^2 a_{jk}^2\right) \\
&= \frac{1}{n^2(n-1)}\sum_{i,j=1,i\neq j}^{n}\sum_{k,l=1}^{n} a_{ik}^2 a_{jl}^2 - \frac{1}{n(n-1)}\sum_{i,j=1,i\neq j}^{n}\sum_{k} a_{ik}^2 a_{jk}^2 \\
&\leq \frac{(n-1)^2}{n^2(n-1)} \\
&\leq \frac{1}{n}
\end{aligned}
$$

It will be convenient to express our bound as a multiple of $\sum_{i,j=1}^{n} a_{i,j}^4$, so we establish a lower bound on that quantity. Our scaling is such that $\sum_{i,j=1}^{n} a_{i,j}^2 = n-1$, so if we write $a := [a_{11}^2\ a_{12}^2 \dots a_{nn}^2]^T$ out as a vector, $a^T 1 = n-1$. By Cauchy-Schwarz,

$$\begin{aligned}(n-1)^2 &= (a^T 1)^2 \\ &\leq ||a||_2^2 ||1||_2^2 \\ &= n^2 \sum_{i,j=1}^{n} a_{i,j}^4.\end{aligned}$$

Therefore, $\sum_{i,j=1}^{n} a_{i,j}^4 \geq 1$, so

$$\sum_{i,j=1, i\neq j}^{n} \mathrm{Cov}(a_{i\Pi(i)}^2, a_{j\Pi(j)}^2) \leq \frac{1}{n} \sum_{i,j=1}^{n} a_{i,j}^4. \tag{B.7}$$

For the second term in (B.6) we again apply corollary A.2:

$$\mathrm{Var}\left(\sum_{i,j=1, i\neq j}^{n} (a_{i\Pi(i)} a_{j\Pi(j)} + a_{i\Pi(j)} a_{j\Pi(i)}) \right) < 2\,\mathrm{Var}\,(X) + 2\,\mathrm{Var}\,(Y)\,,$$

where $X = \sum_{i,j=1, i\neq j}^{n} a_{i\Pi(i)} a_{j\Pi(j)}$ and $Y = \sum_{i,j=1, i\neq j}^{n} a_{i\Pi(j)} a_{j\Pi(i)}$. We note that

$$X = \sum_{i=1}^{n} a_{i\Pi(i)} \sum_{j=1, j\neq i}^{n} a_{j\Pi(j)} = W^2 - \sum_{i=1}^{n} a_{i\Pi(i)}^2. \tag{B.8}$$

TODO: Finish including proof of Hoeffing's Combinatorial Central Limit Theorem. There's still that one bound that I cannot rederive. Well, I can just cite Stein. $\square$

B.0.2 Proof of Lemma 1.5

Proof. Let

$$f(w) = \begin{cases} -3a/2 & w \leq w \leq z - 2a, \\ w - z + a/2 & z - 2a \leq w \leq z + a, \\ 3a/2 & w \geq z + a. \end{cases}$$

Since

$$\mathbb{E}Wf(W) \leq \mathbb{E}[|W||f(W)|] \leq \frac{3a}{2}\mathbb{E}|W| \leq \frac{3a}{2}\sqrt{\mathbb{E}W^2},$$

we have

$$\begin{aligned} 3a\lambda\sqrt{\mathbb{E}W^2} &\geq 2\lambda\mathbb{E}WF(W) \\ &= \mathbb{E}[(W - W')(f(W) - f(W'))] + 2\lambda\mathbb{E}f(W)R \quad \text{by } (A.4) \end{aligned}$$

We also bound the term involving the remainder

$$-2\lambda\mathbb{E}f(W)R \leq 2\lambda\mathbb{E}|f(W)||R| \leq 3a\lambda\mathbb{E}|R|$$

so that

$$\begin{aligned} 3a\lambda(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|) &\geq \mathbb{E}(W - W')(f(W) - f(W')) \\ &= \mathbb{E}\left((W - W')\int_{W'-W}^{0} f'(W + t)dt\right) \\ &\geq \mathbb{E}\left((W - W')\int_{W'-W}^{0} \mathbf{1}_{\{|t|\leq a\}}\mathbf{1}_{\{z-a\leq W\leq z\}}f'(W + t)dt\right). \end{aligned}$$

Since $f'(W + t) = \mathbf{1}_{\{z-2a\leq W+t\leq z+a\}}$,

$$\mathbf{1}_{\{|t|\leq a\}}\mathbf{1}_{\{z-a\leq W\leq z\}}f'(W + t) = \mathbf{1}_{\{|t|\leq a\}}\mathbf{1}_{\{z-a\leq W\leq z\}}.$$

Therefore,

$$
\begin{aligned}
3a\lambda(\sqrt{\mathbb{E}W^2}+\mathbb{E}|R|) &\geq \mathbb{E}\left((W-W')\int_{W'-W}^{0}\mathbf{1}_{\{|t|\leq a\}}dt\mathbf{1}_{\{z-a\leq W\leq z\}}\right)\\
&= \mathbb{E}(|W-W'|\min(a,|W-W'|)\mathbf{1}_{\{z-a\leq W\leq z\}})\\
&\geq \mathbb{E}((W-W')^2\mathbf{1}_{\{0\leq W-W'\leq a\}}\mathbf{1}_{\{z-a\leq W\leq z\}})\\
&= \mathbb{E}((W-W')^2\mathbf{1}_{\{-a\leq W'-W\leq 0\}}\mathbf{1}_{\{z-a\leq W\leq z\}}).
\end{aligned}
$$

The proof of the second claim proceeds similarly:

$$
\begin{aligned}
3a\lambda(\sqrt{\mathbb{E}W^2}+\mathbb{E}|R|) &\geq \mathbb{E}(W-W')(f(W)-f(W'))\\
&= \mathbb{E}(W'-W)(f(W')-f(W))\\
&= \mathbb{E}\left((W'-W)\int_{0}^{W'-W} f'(W+t)dt\right)\\
&\geq \mathbb{E}\left((W'-W)\int_{0}^{W'-W}\mathbf{1}_{\{|t|\leq a\}}\mathbf{1}_{\{z-a\leq W\leq z\}}f'(W+t)dt\right)\\
&= \mathbb{E}\left((W'-W)\int_{0}^{W'-W}\mathbf{1}_{\{|t|\leq a\}}dt\mathbf{1}_{\{z-a\leq W\leq z\}}\right)\\
&= \mathbb{E}(|W'-W|\min(a,|W'-W|)\mathbf{1}_{\{z-a\leq W\leq z\}})\\
&\geq \mathbb{E}((W'-W)^2\mathbf{1}_{\{0\leq W-W'\leq a\}}\mathbf{1}_{\{z-a\leq W\leq z\}}).
\end{aligned}
$$

□

B.0.3 Proof of Theorem 1.6

Proof. For $z\in\mathbb{R}$ and $\alpha>0$ let f be the solution to the Stein equation

$$f'(w)-wf(w)=h_{z,\alpha}(w)-\Phi(z) \tag{B.9}$$

for the smoothed indicator

$$h_{z,\alpha}(w) = \begin{cases} 1 & w \le z \\ 1 + \frac{z-w}{\alpha} & z < w \le z + \alpha \\ 0 & w > z + \alpha. \end{cases} \tag{B.10}$$

Therefore,

$$\begin{aligned} |P(W \le z) - \Phi(z)| &= |\mathbb{E}[(f'(W) - Wf(W))]| \\ &= \left| \mathbb{E}\left[f'(W) - \frac{(W' - W)(f(W') - f(W))}{2\lambda} + f(W)R \right] \right| \\ &= \left| \mathbb{E}\left[f'(W) \left(1 - \frac{(W' - W)^2}{2\lambda} \right) \right.\right. \\ &\quad \left.\left. + \frac{f'(W)(W' - W)^2 - (f(W') - f(W))(W' - W)}{2\lambda} + f(W)R \right] \right| \\ &:= |\mathbb{E}[J_1 + J_2 + J_3]| \\ &\le |\mathbb{E}J_1| + |\mathbb{E}J_2| + |\mathbb{E}J_3|. \end{aligned} \tag{B.11}$$

It is known from Chen and Shao (2004) that for all $w \in \mathbb{R}, 0 \le f(w) \le 1$ and $|f'(w)| \le 1$. Then

$$|\mathbb{E}J_3| \le \mathbb{E}|J_3| = \mathbb{E}|f(W)R| \le \mathbb{E}|R| \tag{B.12}$$

and

$$
\begin{aligned}
|\mathbb{E}J_1| &= \left|\mathbb{E}\left[f'(W)\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right]\right| \\
&\le \mathbb{E}\left[\left|f'(W)\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right|\right] \\
&\le \mathbb{E}\left[\left|\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right|\right] \\
&= \frac{1}{2\lambda}\mathbb{E}[|2\lambda - \mathbb{E}[(W'-W)^2|W]|] \\
&= \frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}W^2 - \mathbb{E}WR) - \mathbb{E}[(W'-W)^2|W] + 2\lambda(1-\mathbb{E}W^2+\mathbb{E}WR)|] \\
&\le \frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}W^2 - \mathbb{E}WR) - \mathbb{E}[(W'-W)^2|W]|] + \mathbb{E}|1-\mathbb{E}W^2+\mathbb{E}WR|
\end{aligned} \tag{B.13}
$$

Note that

$$
\mathbb{E}[\mathbb{E}[(W'-W)^2|W]] = \mathbb{E}\Delta^2 = 2\lambda(\mathbb{E}W^2 - \mathbb{E}WR), \tag{B.14}
$$

so

$$
\frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}W^2 - \mathbb{E}WR) - \mathbb{E}[(W'-W)^2|W]|] \le \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])}. \tag{B.15}
$$

Combining with (B.13),

$$
\begin{aligned}
|\mathbb{E}J_1| &\le \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}|1-\mathbb{E}W^2+\mathbb{E}WR| \\
&\le \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}|1-\mathbb{E}W^2| + \mathbb{E}|WR|
\end{aligned} \tag{B.16}
$$

Lastly, we bound the second term,

$$\begin{aligned}
J_2 &= \frac{1}{2\lambda}(W' - W) \int_W^{W'} (f'(W) - f'(t))dt \\
&= \frac{1}{2\lambda}(W' - W) \int_W^{W'} \int_t^W f''(u)dudt \\
&= \frac{1}{2\lambda}(W' - W) \int_W^{W'} (W' - u)f''(u)du.
\end{aligned} \tag{B.17}$$

To show the final equality, consider separately the cases $W \leq W'$ and $W' \leq W$. For the former,

$$\begin{aligned}
-\frac{1}{2\lambda}(W' - W) \int_W^{W'} \int_W^t f''(u)dudt &= -\frac{1}{2\lambda}(W' - W) \int_W^{W'} \int_u^{W'} f''(u)dtdu \\
&= -\frac{1}{2\lambda}(W' - W) \int_W^{W'} (W' - u)f''(u)du.
\end{aligned}$$

For the latter,

$$\begin{aligned}
\frac{1}{2\lambda}(W' - W) \int_W^{W'} \int_t^W f''(u)dudt &= -\frac{1}{2\lambda}(W' - W) \int_{W'}^{W} \int_t^W f''(u)dudt \\
&= -\frac{1}{2\lambda}(W' - W) \int_{W'}^{W} \int_{W'}^u f''(u)dtdu \\
&= -\frac{1}{2\lambda}(W' - W) \int_{W'}^{W} (u - W')f''(u)du.
\end{aligned}$$

Since W and W' are exchangeable,

$$
\begin{aligned}
|\mathbb{E}J_2| &= \left|\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)\int_W^{W'}(W'-u)f''(u)du\right]\right| \\
&= \left|\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)\int_W^{W'}\left(\frac{W+W'}{2}-u\right)f''(u)du\right]\right| \\
&\leq \left|\mathbb{E}\left[||f''||\frac{1}{2\lambda}|W'-W|\int_{\min(W,W')}^{\max(W,W')}\left|\frac{W+W'}{2}-u\right|du\right]\right| \\
&= \left|\mathbb{E}\left[||f''||\frac{1}{2\lambda}\frac{|W'-W|^3}{4}\right]\right| \\
&\leq \frac{\mathbb{E}|W'-W|^3}{4\alpha\lambda},
\end{aligned}
\tag{B.18}
$$

where the final inequality follows from the fact that $|h'_{z,\alpha}(x)| \leq 1/\alpha$ for all $x \in \mathbb{R}$ and Lemma A.5.

Collecting the bounds, we obtain

$$
\begin{aligned}
P(W \leq z) &\leq \mathbb{E}h_{z,\alpha}(W) \\
&\leq Nh_{z,\alpha} + \frac{\mathbb{E}|W'-W|^3}{4\alpha\lambda} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} \\
&\quad + |1-\mathbb{E}W^2| + \mathbb{E}|WR| + \mathbb{E}|R| \\
&\leq \Phi(z) + \frac{\alpha}{\sqrt{2\pi}} + \frac{\mathbb{E}|W'-W|^3}{4\alpha\lambda} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} \\
&\quad + |\mathbb{E}W^2-1| + \mathbb{E}|WR| + \mathbb{E}|R|
\end{aligned}
\tag{B.19}
$$

The minimizer of the expression is

$$
\alpha = \frac{(2\pi)^{1/4}}{2}\sqrt{\frac{\mathbb{E}|W'-W|^3}{\lambda}}. \tag{B.20}
$$

Plugging this in, we get the upper bound

$$\begin{aligned} P(W \le z) - \Phi(z) \le (2\pi)^{-1/4} \sqrt{\frac{\mathbb{E}|W' - W|^3}{\lambda}} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W' - W)^2|W])} \\ + |\mathbb{E}W^2 - 1| + \mathbb{E}|WR| + \mathbb{E}|R| \end{aligned} \tag{B.21}$$

Proving the corresponding lower bound in a similar manner completes the proof of the theorem. □

B.0.4 Proof of Theorem 1.7

Proof. Now we bound $|\mathbb{E}J_2|$ with $\delta \ge 0$. From (B.11),

$$\begin{aligned} 2\lambda J_2 &= f'(W)(W' - W)^2 - (f(W') - f(W))(W' - W) \\ &= (W' - W) \int_0^{W'-W} (f'(W) - f'(W + t))dt \\ &= (W' - W)\mathbf{1}_{|W'-W|\le\delta} \int_0^{W'-W} (f'(W) - f'(W + t))dt. \end{aligned}$$

Using (A.9), $f'(W) = Wf(W) + \mathbf{1}_{\{w\le z\}} - \Phi(z)$ and $f'(W + t) = (W + t)f(W + t) + \mathbf{1}_{\{w+t\le z\}} - \Phi(z)$. Therefore,

$$\begin{aligned} 2\lambda J_2 = (W' - W)\mathbf{1}_{|W'-W|\le\delta} &\int_0^{W'-W} (Wf(W) - (W + t)f(W + t))dt \\ + (W' - W)\mathbf{1}_{|W'-W|\le\delta} &\int_0^{W'-W} (\mathbf{1}_{\{W\le z\}} - \mathbf{1}_{\{W+t\le z\}})dt \\ \equiv J_{21} + J_{22}. \end{aligned}$$

We apply (A.7) with $w = W$, $u = 0$, and $v = t$ to get

$$
\begin{aligned}
|\mathbb{E}J_{21}| &\le \left|(W' - W)\mathbf{1}_{|W'-W|\le\delta} \int_0^{W'-W} \left(|W| + \frac{\sqrt{2pi}}{4}\right) |t| dt\right| \\
&\le \mathbb{E}\left[\frac{1}{2}|W' - W|^3 \mathbf{1}_{|W'-W|\le\delta} \left(|W| + \frac{\sqrt{2pi}}{4}\right)\right] \\
&\le \frac{1}{2}\delta^3 \left(1 + \frac{\sqrt{2\pi}}{4}\right) \\
&\le .82\delta^3.
\end{aligned}
$$

Now for J_{22}, we consider the two cases according to the sign of $W' - W$. When $W' - W \le 0$, we have

$$
\begin{aligned}
\mathbb{E}J_{22}\mathbf{1}_{\{\delta\le W'-W\le 0\}} &= \mathbb{E}\left[(W' - W)\mathbf{1}_{\{\delta\le W'-W\le 0\}} \int_0^{W'-W} (\mathbf{1}_{\{W\le z\}} - \mathbf{1}_{\{W+t\le z\}}) dt\right] \\
&= \mathbb{E}\left[(W - W')\mathbf{1}_{\{\delta\le W'-W\le 0\}} \int_{W'-W}^{0} (\mathbf{1}_{\{z\le W\le z-t\}}) dt\right] \\
&\le \mathbb{E}\left[(W - W')^2 \mathbf{1}_{\{\delta\le W'-W\le 0\}} \mathbf{1}_{\{z-\delta\le W\le z\}}\right] \\
&\le 3\delta\lambda(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|) \quad \text{by (1.5)}
\end{aligned}
$$

Similarly, when $W' - W > 0$,

$$
\begin{aligned}
\mathbb{E}J_{22}\mathbf{1}_{\{0<W'-W\le\delta\}} &= \mathbb{E}\left[(W' - W)\mathbf{1}_{\{0<W'-W\le\delta\}} \int_0^{W'-W} (\mathbf{1}_{\{W\le z\}} - \mathbf{1}_{\{W+t\le z\}}) dt\right] \\
&= \mathbb{E}\left[(W' - W)\mathbf{1}_{\{0<W'-W\le\delta\}} \int_0^{W'-W} \mathbf{1}_{\{z-t<W\le z\}} dt\right] \\
&\le \mathbb{E}\left[(W' - W)^2 \mathbf{1}_{\{0<W'-W\le\delta\}} \mathbf{1}_{\{z-\delta\le W\le z\}}\right] \\
&\le 3\delta\lambda(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|) \quad \text{by (1.5)}
\end{aligned}
$$

Therefore,

$$\begin{aligned}|\mathbb{E}J_2| &\leq \frac{1}{2\lambda}(|\mathbb{E}J_{21}| + |\mathbb{E}J_{22}|) \\ &\leq \frac{.41\delta^3}{\lambda} + 3\delta(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|).\end{aligned}$$

The result follows from (B.11), noting that J_1 and J_3 stay the same. $\square$

Appendix C

Rate of Convergence Bounds

C.0.5 Proof of Proposition 2.4

Proof.

$$\begin{aligned}
\mathbb{E}T_\Pi^2 &= \frac{N-1}{N}\mathbb{E}\left[\left(\frac{q_\Pi}{d_\Pi}\right)^2\right] && \text{(C.1)} \\
&= \frac{N-1}{N}\mathbb{E}\left[\frac{4N^2\bar{u}_{2,\Pi}^2}{2N-2N\bar{u}_{2,\Pi}^2}\right] \qquad \text{from (2.8)} \\
&= 2(N-1)\mathbb{E}\left[\frac{\bar{u}_{2,\Pi}^2}{1-\bar{u}_{2,\Pi}^2}\right] \\
&= 2(N-1)\mathbb{E}g(\bar{u}_{2,\Pi}), && \text{(C.2)}
\end{aligned}$$

where $g(x) = \frac{x^2}{1-x^2}$. Now we proceed to calculate moments of $\bar{u}_{2,\Pi}$.

Mean-centering the u_i has the effect of mean-centering $\bar{u}_{2,\Pi}$:

$$\mathbb{E}\bar{u}_{2,\Pi} = \frac{1}{N}\mathbb{E}\left[\sum_{i=N+1}^{2N} u_{\Pi(i)}\right] = \frac{1}{N}\sum_{i=N+1}^{2N}\mathbb{E}u_{\Pi(i)} = \frac{1}{N}\sum_{i=N+1}^{2N}\frac{1}{2N}\sum_{j=1}^{2N}u_j = 0$$

Under independence, $\mathrm{Var}(\bar{u}_{2,\Pi})$ would be $\frac{1}{N}$ given the scaling. However, the negative dependence induced by the permutation structure approximately halves this value. The scaling is such that $\mathrm{Var}(u_{\Pi(i)}) = 1$. Under independence and with $i \neq j$,

$\mathrm{Var}(u_{\Pi(i)} + u_{\Pi(j)}) = 2$. Summing only 2 (out of $2N$) values under permutation dependence, $\mathrm{Var}(u_{\Pi(i)} + u_{\Pi(j)}) = 2 - \frac{2}{2N-1}$.

We can't use Serfling's result here because we need more than just an upper bound.

$$
\begin{aligned}
\mathrm{Var}(\bar{u}_{2,\Pi}) &= \frac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=N+1}^{2N} u_{\Pi(i)}\right)^2\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\sum_{i=N+1}^{2N} u_{\Pi(i)}^2 + \sum_{i=N+1}^{2N}\sum_{j=N+1,j\neq i}^{2N} u_{\Pi(i)}u_{\Pi(j)}\right] \\
&= \frac{1}{N^2}\sum_{i=N+1}^{2N}\frac{1}{2N}\sum_{j=1}^{2N} u_j^2 + \frac{1}{N^2}\sum_{i=N+1}^{2N}\sum_{j=N+1,j\neq i}^{2N}\mathbb{E}[u_{\Pi(i)}u_{\Pi(j)}] \\
&= \frac{1}{N} + \frac{1}{N^2}\sum_{i=N+1}^{2N}\sum_{j=N+1,j\neq i}^{2N}\frac{1}{2N}\frac{1}{2N-1}\sum_{k=1}^{2N}\sum_{l=1,l\neq k}^{2N} u_k u_l \\
&= \frac{1}{N} + \frac{1}{N^2}\sum_{i=N+1}^{2N}\sum_{j=N+1,j\neq i}^{2N}\frac{1}{2N}\frac{1}{2N-1}\left(\left(\sum_{k=1}^{2N} u_k\right)^2 - \sum_{k=1}^{2N} u_k^2\right) \\
&= \frac{1}{N} + \frac{1}{N^2}\sum_{i=N+1}^{2N}\sum_{j=N+1,j\neq i}^{2N}\frac{1}{2N}\frac{1}{2N-1}\left(0^2 - 2N\right) \\
&= \frac{1}{N} + \frac{1}{N}(N^2 - N)\left(-\frac{1}{2N-1}\right) \\
&= \frac{2N-1}{N(2N-1)} + \frac{1-N}{N(2N-1)} \\
&= \frac{1}{2N-1}
\end{aligned}
$$

Having established the first two moments, we compute the third degree Taylor expansion and bound the error in the approximation. By Taylor's theorem, we expand the function $g(\bar{u}_{2,\Pi}) = \frac{\bar{u}_{2,\Pi}^2}{1-\bar{u}_{2,\Pi}^2}$ around $\mathbb{E}[\bar{u}_{2,\Pi}] = 0$:

$$
g(\bar{u}_{2,\Pi}) = \frac{\bar{u}_{2,\Pi}^2}{1-\bar{u}_{2,\Pi}^2} = g(0) + g'(0)\bar{u}_{2,\Pi} + \frac{g''(0)}{2!}\bar{u}_{2,\Pi}^2 + \frac{g^{(3)}(0)}{3!}\bar{u}_{2,\Pi}^3 + R_3(\bar{u}_{2,\Pi}),
$$

where $R_3(\bar{u}_{2,\Pi}) = \frac{g^{(4)}(\xi_L)}{4!}\bar{u}_{2,\Pi}^4$, with $\xi_L \in [0, \bar{u}_{2,\Pi}]$.

From (C.2) and evaluating the Taylor series, we have

$$\mathbb{E}g(\bar{u}_{2,\Pi}) = \frac{\mathbb{E}T_\Pi^2}{2(N-1)} = \mathbb{E}[\bar{u}_{2,\Pi}^2 + R_3(\bar{u}_{2,\Pi})].$$

Therefore,

$$\begin{aligned}
\left|\frac{\mathbb{E}T_\Pi^2}{2(N-1)} - \mathbb{E}\bar{u}_{2,\Pi}^2\right| &= \left|\frac{\mathbb{E}T_\Pi^2}{2(N-1)} - \frac{1}{2N-1}\right| \\
&\leq \mathbb{E}|R_3(\bar{u}_{2,\Pi})| \\
&= \mathbb{E}\left|\frac{24(5\xi_L^4 + 10\xi_L^2 + 1)}{4!\,(\xi_L - 1)^5}\bar{u}_{2,\Pi}^4\right| \\
&\leq \mathbb{E}\left|\frac{24(5\bar{u}_{2,\Pi}^4 + 10\bar{u}_{2,\Pi}^2 + 1)}{4!\,(\bar{u}_{2,\Pi} - 1)^5}\bar{u}_{2,\Pi}^4\right| \\
&\leq \frac{5B^4 + 10B^2 + 1}{|B-1|^5}\mathbb{E}\bar{u}_{2,\Pi}^4 \\
&\leq \frac{5B^4 + 10B^2 + 1}{|B-1|^5} f_{c_1}(4)N^{-2} \qquad \text{by (2.4)} \\
&:= c_1 N^{-2}
\end{aligned}$$

$$\begin{aligned}
|\mathbb{E}T_\Pi^2 - 1| - \frac{1}{2N-1} &\leq \left|\mathbb{E}T_\Pi^2 - 1 + \frac{1}{2N-1}\right| \\
&= \left|\mathbb{E}T_\Pi^2 - \frac{2(N-1)}{2N-1}\right| \\
&= 2(N-1)\left|\frac{\mathbb{E}T_\Pi^2}{2(N-1)} - \frac{1}{2N-1}\right| \\
&\leq c_1 2(N-1)N^{-2}
\end{aligned}$$

This implies that

$$|\mathbb{E}T_\Pi^2 - 1| \leq \frac{1}{2N-1} + c_1\frac{2N-2}{N^2} \leq \frac{1+2c_1}{N} := c_2 N^{-1}$$

□

C.0.6 Proof of Proposition 2.3

Proof. With two applications of the c_r inequality, we can bound the variance of the sum by a constant times the sum of the variances. Suppose X, Y, and Z have finite variances. Then, with the centered random variables represented by $\tilde{X}, \tilde{Y}$, and $\tilde{Z}$, we have that

$$\begin{aligned}
\mathrm{Var}(X+Y+Z) &= \mathrm{Var}(\tilde{X}+\tilde{Y}+\tilde{Z}) \\
&= \mathbb{E}|(\tilde{X}+\tilde{Y})+\tilde{Z}|^2 \\
&\leq 2\mathbb{E}|\tilde{X}+\tilde{Y}|^2 + 2\mathbb{E}|\tilde{Z}|^2 \\
&\leq 2(2\mathbb{E}\tilde{X}^2 + 2\mathbb{E}\tilde{Y}^2) + 2\mathbb{E}\tilde{Z}^2 \\
&\leq 4(\mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z))
\end{aligned}$$

From (2.11),

$$\begin{aligned}
\mathrm{Var}(\mathbb{E}[(T'_\Pi - T_\Pi)^2|\Pi=\pi]) &= \mathrm{Var}\left(\frac{N-1}{N}\mathbb{E}\left[\left(\frac{2u_{\Pi(J)}-2u_{\Pi(I)}}{d_\Pi} + T'_\Pi\frac{d_\Pi - d'_\Pi}{d_\Pi}\right)^2\middle|\Pi=\pi\right]\right) \\
&\leq \mathrm{Var}\left(\mathbb{E}\left[\left(\frac{2u_{\Pi(J)}-2u_{\Pi(I)}}{d_\Pi} + T'_\Pi\frac{d_\Pi - d'_\Pi}{d_\Pi}\right)^2\middle|\Pi=\pi\right]\right) \\
&\leq 4(\mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z))
\end{aligned}$$

where

$$\begin{aligned}
X &= \mathbb{E}\left[\left(\frac{2u_{\Pi(J)}-2u_{\Pi(I)}}{d_\Pi}\right)^2\middle|\Pi=\pi\right] \\
Y &= \mathbb{E}\left[\left(T'_\Pi\frac{d_\Pi - d'_\Pi}{d_\Pi}\right)^2\middle|\Pi=\pi\right] \\
Z &= 2\mathbb{E}\left[\left(\frac{2u_{\Pi(J)}-2u_{\Pi(I)}}{d_\Pi}T'_\Pi\frac{d_\Pi - d'_\Pi}{d_\Pi}\right)\middle|\Pi=\pi\right]
\end{aligned}$$

The X term will dominate, so we can afford to use coarser methods on Y and Z.

The $\mathbb{E}[u_{\Pi(J)} - u_{\Pi(I)}|\Pi=\pi]$ term is common to applications of Stein's method of exchangeable pairs. However, there is a complication in the d_Π random variable in the denominator. Our strategy will be to calculate the two variances separately with some necessary additional terms.

First, we prove an intermediate result regarding the variance of a product of random variables

$$W = (d_\Pi)^{-2} \text{ and } V = \mathbb{E}[(u_{\Pi(J)} - u_{\Pi(I)})^2 | \Pi = \pi].$$

Then $\mathrm{Var}(X) = 4\,\mathrm{Var}(WV)$ since d_Π is $\sigma(\Pi)$-measurable and

$$\begin{aligned}
\mathrm{Var}(WV) &= \mathrm{Var}(W(V - \mathbb{E}V) + W\mathbb{E}V) \\
&\le 2\,\mathrm{Var}(W(V - \mathbb{E}V)) + 2\,\mathrm{Var}(W\mathbb{E}V) \\
&\le 2\mathbb{E}[W^2(V - \mathbb{E}V)^2] + 2(\mathbb{E}V)^2\,\mathrm{Var}(W) \\
&\le 2(f_{c_2}(2))^2 N^{-2}\,\mathrm{Var}(V) + 2u_\Delta^4\,\mathrm{Var}(W).
\end{aligned} \tag{C.3}$$

$$\begin{aligned}
\mathrm{Var}(W) &= \mathrm{Var}((d_\Pi)^{-2}) \\
&= \mathrm{Var}\left(\frac{1}{2N(1-\bar{u}_{2,\Pi}^2)}\right) \\
&= \frac{1}{4N^2}\left[\mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2\right] \\
&= \frac{1}{4N^2}[\mathbb{E}h(\bar{u}_{2,\Pi}) - (\mathbb{E}\tilde{h}(\bar{u}_{2,\Pi}))^2],
\end{aligned}$$

where

$$h(x) = \left(\frac{1}{1-x^2}\right)^2 = 1 + 2x^2 + 3x^4 + \ldots \text{ and } \tilde{h}(x) = \frac{1}{1-x^2} = 1 + x^2 + x^4 + \ldots$$

By Taylor's theorem,

$$\mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right] = 1 + 2\left(\frac{1}{2N-1}\right) + \mathbb{E}R_3(\bar{u}_{2,\Pi}),$$

with

$$\mathbb{E}|R_3(\bar{u}_{2,\Pi})| \leq \frac{24(35B^4+42B^2+3)}{4!\,(B-1)^6} f_{c_1}(4)N^{-2} := c_4 N^{-2}$$

Re-arranging, we get

$$\left|\mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right] - 1 - \frac{2}{2N-1}\right| \leq c_4 N^{-2}.$$

Applying Taylor's theorem to $\tilde{h}$:

$$\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right] = 1 + \frac{1}{2N-1} + \mathbb{E}\tilde{R}_3(\bar{u}_{2,\Pi}),$$

with

$$\mathbb{E}|\tilde{R}_3(\bar{u}_{2,\Pi})| \leq \frac{24(5B^4+10B^2+1)}{4!\,(B-1)^5} f_{c_1}(4)N^{-2} := c_5 N^{-2}$$

Squaring, applying the bound, and re-arranging yields

$$\left|\left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2 - \left(1+\frac{1}{2N-1}\right)^2\right| \leq 2\left(1+\frac{1}{2N-1}\right)c_5 N^{-2} + c_5^2 N^{-4}$$

Now we combine bounds to get

$$
\begin{aligned}
&\left|\mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right]-\left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2\right| \\
&=\left|\mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right]-\left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2+\frac{1}{(2N-1)^2}-\frac{1}{(2N-1)^2}\right| \\
&\leq\left|\mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right]-\left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2+\frac{1}{(2N-1)^2}\right|+\left|\frac{1}{(2N-1)^2}\right| \\
&\leq\left|\mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right]-1-\frac{2}{2N-1}-\left(\left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2-\left(1+\frac{1}{2N-1}\right)^2\right)\right|+\left|\frac{1}{(2N-1)^2}\right| \\
&\leq c_4N^{-2}+2\left(1+\frac{1}{2N-1}\right)c_5N^{-2}+c_5^2N^{-4}+\left|\frac{1}{(2N-1)^2}\right| \\
&\leq (c_4+3c_5+c_5^2+\frac{1}{4})N^{-2} \\
&:= c_6N^{-2}
\end{aligned}
$$

Therefore, $\mathrm{Var}(W)\leq \frac{c_6}{4}N^{-4}$ and

$$
\mathrm{Var}(X)\leq 8(f_{c_2}(2))^2N^{-2}\,\mathrm{Var}(V)+8u_\Delta^4\frac{c_6}{4}N^{-4}
$$

with

$$
\begin{aligned}
\mathrm{Var}(V) &= \mathrm{Var}(\mathbb{E}[(u_{\Pi(J)}-u_{\Pi(I)})^2|\Pi=\pi]) \\
&= \mathrm{Var}(\mathbb{E}[u_{\Pi(J)}^2+u_{\Pi(I)}^2-2u_{\Pi(J)}u_{\Pi(I)}|\Pi=\pi]) \\
&= \mathrm{Var}\left(\frac{1}{N^2}\sum_{I=1}^{N}\sum_{J=N+1}^{2N}(u_{\pi(J)}^2+u_{\pi(I)}^2-2u_{\pi(J)}u_{\pi(I)})\right) \\
&= \mathrm{Var}\left(\frac{1}{N^2}\left(N\sum_{K=1}^{2N}u_K^2-\sum_{I=1}^{N}\sum_{J=N+1}^{2N}2u_{\pi(J)}u_{\pi(I)}\right)\right) \\
&= \frac{4}{N^4}\sum_{I=1}^{N}\sum_{J=N+1}^{2N}\sum_{K=1}^{N}\sum_{L=N+1}^{2N}\mathrm{Cov}(u_{\pi(I)}u_{\pi(J)},u_{\pi(K)}u_{\pi(L)})
\end{aligned}
$$

since $\sum_{K=1}^{2N} u_K^2 = 2N$ is a constant. We proceed by calculating

$$\mathrm{Cov}(u_{\pi(I)}u_{\pi(J)}, u_{\pi(K)}u_{\pi(L)}) = \mathbb{E}[u_{\pi(I)}u_{\pi(J)}u_{\pi(K)}u_{\pi(L)}] - \mathbb{E}[u_{\pi(I)}u_{\pi(J)}]\mathbb{E}[u_{\pi(K)}u_{\pi(L)}].$$

The index sets for variables I and J (and K and L) are disjoint, so

$$\mathbb{E}[u_{\pi(I)}u_{\pi(J)}] = \mathbb{E}[u_{\pi(K)}u_{\pi(L)}] = \frac{1}{2N}\frac{1}{2N-1}\sum_{I=1}^{2N} u_I \sum_{J=1,J\neq I}^{2N} u_J = -\frac{1}{2N-1}$$

for all values of I, J, K, L in the sum. Therefore,

$$\mathbb{E}[u_{\pi(I)}u_{\pi(J)}] = \mathbb{E}[u_{\pi(K)}u_{\pi(L)}] = \frac{1}{(2N-1)^2}.$$

However, K could equal I and L could equal J, which changes the mass assigned by the permutation distribution, necessitating a separate treatment for each case.

Case $I \neq J \neq K \neq L$:

$$\begin{aligned}
&\mathbb{E}[u_{\pi(I)}u_{\pi(J)}u_{\pi(K)}u_{\pi(L)}] \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\frac{1}{2N-3}\sum_{I=1}^{2N}\sum_{J=1,J\neq I}^{2N}\sum_{K=1,K\neq I,J}^{2N}\sum_{L=1,L\neq I,J,K}^{2N} u_I u_J u_K u_L \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\frac{1}{2N-3}\sum_{I=1}^{2N} u_I \sum_{J=1,J\neq I}^{2N} u_J \sum_{K=1,K\neq I,J}^{2N} u_K(-u_I - u_J - u_K) \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\frac{1}{2N-3}\sum_{I=1}^{2N} u_I \sum_{J=1,J\neq I}^{2N} u_J((-u_I - u_J)(-u_I - u_J) + (u_I^2 + u_J^2 - 2N)) \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\frac{1}{2N-3}\sum_{I=1}^{2N} u_I \sum_{J=1,J\neq I}^{2N} u_J(2u_I^2 - 2N + 2u_J^2 + 2u_I u_J) \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\frac{1}{2N-3}\sum_{I=1}^{2N} u_I \left((2u_I^2 - 2N)(-u_I) + 2\sum_{J=1,J\neq I}^{2N} u_J^3 + 2u_I(2N - u_I^2)\right) \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\frac{1}{2N-3}\sum_{I=1}^{2N} u_I \left(-4u_I^3 + 6Nu_I + 2\left(\sum_{J=1}^{2N} u_J^3 - u_I^3\right)\right) \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\frac{1}{2N-3}\left(-6\sum_{I=1}^{2N} u_I^4 + 12N^2\right)
\end{aligned}$$

for $N^2(N-1)^2$ terms in the sum.

Case $I=K$ and $J=L$:

$$\begin{aligned}
\mathbb{E}[u^2_{\pi(I)}u^2_{\pi(J)}] &= \frac{1}{2N}\frac{1}{2N-1}\sum_{I=1}^{2N}\sum_{J=1,J\neq I}^{2N} u_I^2 u_J^2 \\
&= \frac{1}{2N}\frac{1}{2N-1}\sum_{I=1}^{2N} u_I^2(2N-u_I^2) \\
&= \frac{2N}{2N-1} - \frac{1}{2N}\frac{1}{2N-1}\sum_{I=1}^{2N} u_I^4
\end{aligned}$$

for N^2 terms in the sum.

Case $I=K, J\neq L$ or $I\neq K, J=L$:

$$\begin{aligned}
\mathbb{E}[u^2_{\pi(I)}u_{\pi(J)}u_{\pi(K)}] &= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\sum_{I=1}^{2N}\sum_{J=1,J\neq I}^{2N}\sum_{K=1,K\neq I,J}^{2N} u_I^2 u_J u_K \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\sum_{I=1}^{2N}\sum_{J=1,J\neq I}^{2N} u_I^2 u_J(0-u_I-u_J) \\
&= -\frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\left(\sum_{I=1}^{2N} u_I^3 \sum_{J=1,J\neq I}^{2N} u_J + \sum_{I=1}^{2N} u_I^2 \sum_{J=1,J\neq I}^{2N} u_J^2\right) \\
&= -\frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\left(\sum_{I=1}^{2N} -u_I^4 + \sum_{I=1}^{2N} u_I^2(2N-u_I^2)\right) \\
&= \frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\left(2\sum_{I=1}^{2N} u_I^4 - 4N^2\right)
\end{aligned}$$

for $2N^2(N-1)$ terms in the sum.

Putting it all together, we have

$$
\begin{aligned}
&\mathrm{Var}(\mathbb{E}[(u_{\Pi(J)} - u_{\Pi(i)})^2]|\Pi = \pi) \\
&= \frac{4}{N^4}(N^2(N-1)^2)\left(\frac{1}{(2N)(2N-1)(2N-2)(2N-3)}\left(-6\sum_{i=1}^{2N} u_i^4 + 12N^2\right) - \frac{1}{(2N-1)^2}\right) \\
&+ \frac{4}{N^4}N^2\left(\frac{2N}{2N-1} - \frac{1}{2N}\frac{1}{2N-1}\sum_{i=1}^{2N} u_i^4 - \frac{1}{(2N-1)^2}\right) \\
&+ \frac{4}{N^4}(2N^2(N-1))\left(\frac{1}{2N}\frac{1}{2N-1}\frac{1}{2N-2}\left(2\sum_{i=1}^{2N} u_i^4 - 4N^2\right) - \frac{1}{(2N-1)^2}\right) \\
&\leq \frac{48}{4N^2} + \frac{8}{N^2} + \frac{16\sum_{i=1}^{2N} u_i^4}{N^4} \\
&= \left(20 + 16\left(\sum_{i=1}^{2N} u_i^4\right)N^{-2}\right)N^{-2}
\end{aligned}
$$

Therefore,

$$
\mathrm{Var}(X) \leq 8(f_{c_2}(2))^2\left(20 + 16\left(\sum_{i=1}^{2N} u_i^4\right)N^{-2}\right)N^{-4} + 8u_\Delta^4\frac{c_6}{4}N^{-4}
$$

Because the latter two terms are much smaller in order, we can apply coarser techniques. In particular, we use the following bound:

$$
\mathrm{Var}(\mathbb{E}[U|V]) = \mathrm{Var}(U) - \mathbb{E}(\mathrm{Var}(U|V)) \leq E[U^2]
$$

Applying to the second term,

$$\begin{aligned}
\operatorname{Var}(Y) &= \operatorname{Var}\left(\mathbb{E}\left[\left(T'_\Pi \frac{d_\Pi - d'_\Pi}{d_\Pi}\right)^2 \middle| \Pi = \pi\right]\right) \\
&\leq \mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}(d_\Pi - d'_\Pi)\right)^4\right] \\
&\leq \sqrt{\mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^8\right] \mathbb{E}[(d_\Pi - d'_\Pi)^8]} \\
&\leq \sqrt{f_{c_6}(8) N^{-8/2} f_{c_4}(8) N^{-8}} \text{ from (2.10), (2.7)} \\
&= \sqrt{f_{c_6}(8) f_{c_4}(8)} N^{-6} \\
&:= c_7 N^{-6}
\end{aligned}$$

And to the third,

$$\begin{aligned}
\operatorname{Var}(Z) &= 4 \operatorname{Var}\left(\mathbb{E}\left[\left(\frac{2u_{\Pi(J)} - 2u_{\Pi(I)}}{d_\Pi} T'_\Pi \frac{d_\Pi - d'_\Pi}{d_\Pi}\right) \middle| \Pi = \pi\right]\right) \\
&\leq 16 u_\Delta^2 \mathbb{E}\left[\left(\frac{1}{d_\Pi} \frac{q'_\Pi}{d_\Pi d'_\Pi}(d_\Pi - d'_\Pi)\right)^2\right] \\
&\leq 16 u_\Delta^2 f_{c_2}(2) N^{-2/2} \sqrt{\mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^4\right] \mathbb{E}[(d_\Pi - d'_\Pi)^4]} \text{ from (2.5)} \\
&\leq 16 u_\Delta^2 f_{c_2}(2) N^{-1} \sqrt{f_{c_6}(4) N^{-4/2} f_{c_4}(4) N^{-4}} \text{ from (2.10), (2.7)} \\
&\leq 16 u_\Delta^2 f_{c_2}(2) (f_{c_6}(4))^{-1/2} (f_{c_4}(4))^{-1/2} N^{-4} \\
&:= c_8 N^{-4}
\end{aligned}$$

$$
\begin{aligned}
&\frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(T'_\Pi - T_\Pi)^2|T_\Pi])} \\
&= N\sqrt{(\mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z))} \\
&\leq N\sqrt{8(f_{c_2}(2))^2\left(20 + 16\left(\sum_{i=1}^{2N} u_i^4\right)N^{-2}\right)N^{-4} + 8u_\Delta^4\frac{c_6}{4}N^{-4} + c_7N^{-6} + c_8N^{-4}} \\
&:= N^{-1}c_3\sqrt{20 + 16\frac{\sum_{i=1}^{2N} u_i^4}{N^2}}
\end{aligned}
$$

□

C.0.7 Proof of Proposition 2.2

Proof. The strategy is to break apart the remainder term from the main piece. From (2.11),

$$
\begin{aligned}
\mathbb{E}|T'_\Pi - T_\Pi|^3 &= \left(\frac{N-1}{N}\right)^{3/2}\mathbb{E}\left[d_\Pi^{-3}\left|2u_{\Pi(J)} - 2u_{\Pi(I)} + q'_\Pi\frac{d_\Pi - d'_\Pi}{d'_\Pi}\right|^3\right] \\
&\leq 8\left(8u_\Delta^3\mathbb{E}[d_\Pi^{-3}] + \sqrt{\mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^6\right]\mathbb{E}[(d_\Pi - d'_\Pi)^6]}\right) \\
&\leq 64u_\Delta^3 f_{c_2}(3)N^{-3/2} + 8\sqrt{f_{c_6}(6)N^{-6/2}f_{c_4}(6)N^{-6}} \text{ from } (2.5), (2.10), (2.7) \\
&\leq \frac{c_9^2}{2}N^{-3/2}
\end{aligned}
$$

Therefore,

$$
(2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|T'_\Pi - T_\Pi|^3}{\lambda}} \leq (2\pi)^{-1/4}c_9N^{-1/4}.
$$

□

C.0.8 Proof of Proposition 2.6

Proof.

$$\begin{aligned}
\mathbb{E}|R| &= \mathbb{E}\left|\binom{N}{2}\sqrt{\frac{N-1}{N}}\frac{1}{d_\Pi}\mathbb{E}\left[q'_\Pi\frac{(d_\Pi - d'_\Pi)}{d'_\Pi}\middle|T_\Pi\right]\right| \\
&\le \frac{N}{2}\mathbb{E}\left|\frac{q'_\Pi}{d_\Pi d'_\Pi}(d_\Pi - d'_\Pi)\right| \\
&\le \frac{N}{2}\sqrt{\mathbb{E}\left|\frac{q'_\Pi}{d_\Pi d'_\Pi}\right|^2\mathbb{E}[d_\Pi - d'_\Pi]^2} \\
&\le \frac{N}{2}\sqrt{f_{c_6}(2)N^{-2/2}f_{c_4}(2)N^{-2}} \text{ from } (2.10), (2.7) \\
&= \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}
\end{aligned}$$

□

C.0.9 Proof of Proposition 2.5

Proof.

$$\begin{aligned}
\mathbb{E}|T_\Pi R| &= \mathbb{E}\left|T_\Pi\binom{N}{2}\sqrt{\frac{N-1}{N}}\frac{1}{d_\Pi}\mathbb{E}\left[q'_\Pi\frac{(d_\Pi - d'_\Pi)}{d'_\Pi}\middle|T_\Pi\right]\right| \\
&\le \frac{N}{2}\mathbb{E}\left|T_\Pi\frac{q'_\Pi}{d_\Pi d'_\Pi}(d_\Pi - d'_\Pi)\right| \\
&\le \frac{N}{2}\sqrt{\mathbb{E}T_\Pi^2\mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^2(d_\Pi - d'_\Pi)^2\right]} \\
&\le \frac{N}{2}\sqrt{\mathbb{E}T_\Pi^2\sqrt{\mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^4\right]\mathbb{E}[(d_\Pi - d'_\Pi)^4]}} \\
&\le \frac{N}{2}\sqrt{\mathbb{E}T_\Pi^2\sqrt{f_{c_6}(4)N^{-4/2}f_{c_4}(4)N^{-4}}} \text{ from } (2.10), (2.7) \\
&= \frac{N^{-1/2}}{2}(f_{c_6}(4)f_{c_4}(4))^{1/4}\sqrt{\mathbb{E}T_\Pi^2} \\
&\le \frac{1}{2}(f_{c_6}(4)f_{c_4}(4))^{1/4}\sqrt{2+2c_1}N^{-1/2}
\end{aligned}$$

because $\mathbb{E}T_\Pi^2 \leq 1 + \frac{1+2c_1}{N} \leq 2 + 2c_1$. $\square$

References

[1] V. Bentkus and F. Götze. The berry-esseen bound for student's statistic. *The Annals of Probability*, 24(1):491–503, 1996.

[2] Erwin Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Probability Theory and Related Fields*, 66(3):379–386, 1984.

[3] L.H.Y. Chen, L. Goldstein, and Q.M. Shao. *Normal Approximation by Stein's Method*. Springer Verlag, 2010.

[4] H.A. David. The beginnings of randomization tests. *The American Statistician*, 62(1):70–72, 2008.

[5] P. Diaconis and S. Holmes. Gray codes for randomization procedures. *Statistics and Computing*, 4(4):287–302, 1994.

[6] P. Diaconis and E. Lehmann. Comment. *Journal of the American Statistical Association*, 103(481):16–19, 2008.

[7] R.A. Fisher. *The design of experiments.* Oliver & Boyd, 1935.

[8] E.L. Lehmann. *Elements of large-sample theory.* Springer Verlag, 1999.

[9] J. Ludbrook and H. Dudley. Why permutation tests are superior to t and f tests in biomedical research. *American Statistician*, pages 127–132, 1998.

[10] J. Neyman and E.S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. *Biometrika*, 20(1/2):175–240, 1928.

[11] R.J. Serfling. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, 2(1):39–48, 1974.

[12] Q.M. Shao. An explicit berry-esseen bound for student's t-statistic via stein's method. *Stein's Method and Applications (AD Barbour and LHY Chen eds). Lecture Notes Series, Institute for Mathematical Sciences, NUS*, 5:143–155, 2005.

[13] C. Stein. Approximate computation of expectations. *Lecture Notes-Monograph Series*, 7, 1986.

[14] SL Zabell. On Student's 1908 Article The Probable Error of a Mean. *Journal of the American Statistical Association*, 103(481):1–7, 2008.