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 for the randomization distribution of the t-statistic.

We also include other results that will be useful for our proof.

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. We reproduce Stein's [32] proof of the Hoeffding combinatorial central limit theorem (HCCLT) with explicit calculation of various constants. 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 an r.v. in the exchangeable pair.

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

Theorem 1.1 (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*

$$\mathbb{E}|X+Y|^r < c_r(\mathbb{E}|X|^r + \mathbb{E}|Y|^r), \tag{1.1}$$

where $c_r = 1$ *when* $r \le 1$ *and* $c_r = 2^{r-1}$ *when* $r \ge 1$.

Corollary 1.2. *Suppose that* $\mathrm{Var}(X) < \infty$ *and* $\mathrm{Var}(Y) < \infty$. *Then*

$$\mathrm{Var}(X+Y) < 2(\mathrm{Var}(X) + \mathrm{Var}(Y)). \tag{1.2}$$

Proof. This follows immediately by applying Theorem 1.1 to the centered random variables $X' = X - \mathbb{E}X$ and $Y' = Y - \mathbb{E}Y$. □

1.2 Hoeffding's Combinatorial CLT

Theorem 1.3. *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* $n-1$*:*

$$\sum_{j=1}^{n} a_{ij} = 0 \tag{1.3}$$

$$\sum_{i=1}^{n} a_{ij} = 0 \tag{1.4}$$

$$\sum_{i=1,j=1}^{n} a_{ij}^2 = n-1 \tag{1.5}$$

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

$$P(\Pi = \pi) = \frac{1}{n!}. \tag{1.6}$$

Define

$$W = \sum_{i=1}^{n} a_{i\Pi(i)} \tag{1.7}$$

to be the sum of a random diagonal. Then

$$|P(W \leq w) - \Phi(w)| \leq \frac{C}{\sqrt{n}} \left[\sqrt{\sum_{i,j=1}^{n} a_{ij}^4} + \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^3} \right]. \tag{1.8}$$

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:

$$P(I = i, J = j) = \frac{1}{n(n-1)} \quad i, j \in \{1, \ldots, n\}, i \neq j. \tag{1.9}$$

Define the random permutation Π' by

$$\Pi'(i) = \Pi \circ (I, J) = \begin{cases} \Pi(J) & i = I \\ \Pi(I) & i = J \\ \Pi(i) & \text{else.} \end{cases} \tag{1.10}$$

We construct our exchangeable pair by defining

$$W' = \sum_{i=1}^{n} a_{i\Pi'(i)} = W - a_{I\Pi(I)} + a_{I\Pi(J)} - a_{J\Pi(J)} + a_{J\Pi(I)}. \tag{1.11}$$

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

$$\lambda = \frac{2}{n-1}. \tag{1.12}$$

To bound the variance component, compute

$$
\begin{aligned}
\mathbb{E}[(W-W')^2|\Pi] &= \mathbb{E}[(a_{I\Pi(I)} - a_{I\Pi(J)} + a_{J\Pi(J)} - a_{J\Pi(I)})^2|\Pi] \\
&= \mathbb{E}[a^2_{I\Pi(I)} + a^2_{J\Pi(J)} + a^2_{I\Pi(J)} + a^2_{J\Pi(I)} \\
&\quad - 2a_{I\Pi(I)}a_{I\Pi(J)} - 2a_{J\Pi(J)}a_{J\Pi(I)} - 2a_{I\Pi(I)}a_{J\Pi(I)} - 2a_{J\Pi(J)}a_{I\Pi(J)} \\
&\quad + 2a_{I\Pi(I)}a_{J\Pi(J)} + 2a_{I\Pi(J)}a_{J\Pi(I)}|\Pi] \\
&= \frac{2}{n}\sum_{i=1}^n a^2_{i\Pi(i)} + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^n a^2_{i\Pi(j)} \\
&\quad - \frac{4}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^n a_{i\Pi(i)}a_{i\Pi(j)} - \frac{4}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^n a_{i\Pi(i)}a_{j\Pi(i)} \\
&\quad + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^n a_{i\Pi(i)}a_{j\Pi(j)} + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^n a_{i\Pi(j)}a_{j\Pi(i)} \\
&= \frac{2}{n}\sum_{i=1}^n a^2_{i\Pi(i)} + \frac{2}{n}\frac{1}{n-1}\left(\sum_{i,j=1}^n a^2_{i\Pi(j)} - \sum_{i=1}^n a^2_{i\Pi(i)}\right) \\
&\quad - \frac{4}{n}\frac{1}{n-1}\sum_{i=1}^n \left(a_{i\Pi(i)}\sum_{j=1}^n \left(a_{i\Pi(j)} + a_{j\Pi(i)}\right) - 2a^2_{i\Pi(i)}\right) \\
&\quad + \frac{2}{n}\frac{1}{n-1}\left(\sum_{i,j=1,i\neq j}^n a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right) \\
&= \frac{2}{n}\left(1 - \frac{1}{n-1}\right)\sum_{i=1}^n a^2_{i\Pi(i)} + \frac{2}{n} \\
&\quad + \frac{8}{n}\frac{1}{n-1}\sum_{i=1}^n a^2_{i\Pi(i)} \\
&\quad + \frac{2}{n}\frac{1}{n-1}\sum_{i=1}^n\sum_{j=1}^n \left(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right) - \frac{4}{n}\frac{1}{n-1}\sum_{i=1}^n a^2_{i\Pi(i)} \\
&= \frac{2}{n} + \frac{2(n+2)}{n(n-1)}\sum_{i=1}^n a^2_{i\Pi(i)} + \frac{2}{n(n-1)}\sum_{i,j=1,i\neq j}^n (a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)})
\end{aligned}
\tag{1.13}
$$

From (1.13) and corollary 1.2,

$$\begin{aligned}
\mathbb{E}[(W-W')^2|\Pi] &= \operatorname{Var}\left(\frac{2(n+2)}{n(n-1)}\sum_{i=1}^n a_{i\Pi(i)}^2\right. \\
&\quad \left. + \frac{2}{n(n-1)}\sum_{i,j=1,i\neq j}^n \left(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right)\right) \\
&\le 2\left(\frac{4(n+2)^2}{n^2(n-1)^2}\operatorname{Var}\left(\sum_{i=1}^n a_{i\Pi(i)}^2\right) + \right. \\
&\quad \left. \frac{4}{n^2(n-1)^2}\operatorname{Var}\left(\sum_{i,j=1,i\neq j}^n \left(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right)\right)\right) \\
&\le \frac{32}{n^2}\operatorname{Var}\left(\sum_{i=1}^n a_{i\Pi(i)}^2\right) + \frac{32}{n^4}\operatorname{Var}\left(\sum_{i,j=1,i\neq j}^n \left(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right)\right)
\end{aligned} \tag{1.14}$$

for $n \ge 2$ since $n-1 \ge n/2 \implies \frac{1}{(n-1)^2} \le \frac{4}{n^2}$ for $n \ge 2$.

First, we address the first term in (1.14):

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

with

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

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 $\mathbf{a} := [a_{11}^2\ a_{12}^2 \ldots a_{nn}^2]^T$ out as a vector, $\mathbf{a}^T\mathbf{1} = n-1$. By Cauchy-Schwarz,

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

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

$$\sum_{i,j=1,i\neq j}^{n} \mathrm{Cov}(a_{i\Pi(i)}^2, a_{j\Pi(j)}^2) \le \frac{1}{n} \sum_{i,j=1}^{n} a_{i,j}^4. \tag{1.15}$$

For the second term in (1.14) we again apply corollary 1.2:

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

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

$$X = \sum_{i=1}^{n} a_{i\Pi(i)} \sum_{j=1,j\neq i}^{n} a_{j\Pi(j)} = W^2 - \sum_{i=1}^{n} a_{i\Pi(i)}^2. \tag{1.16}$$

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. □

1.3 Exchangeable Pairs

TODO: Add a lot of development for exchangeable pairs. For now, focusing on generalizing the theorems in "Normal Approximation by Stein's Method." [4]

Theorem 5.5 in "Normal Approximation by Stein's Method" concerns variance

1 exchangeable random variables. Our setting has the variance tending to 1, so we first prove a slight generalization of the theorem. Large parts of the proof are copied verbatim from the book.

1.4 Preliminaries

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

$$\mathbb{E}[W - W'|W] = \lambda(W - R), \tag{1.17}$$

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

Lemma 1.5. *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 [4]:

Lemma 1.6. *Let (W, W') be an approximate Stein pair and $\Delta = W - W'$. Then*

$$\mathbb{E}W = \mathbb{E}R \quad \text{and} \quad \mathbb{E}\Delta^2 = 2\lambda\mathbb{E}W^2 - 2\lambda\mathbb{E}WR \quad \text{if } \mathbb{E}W^2 < \infty. \tag{1.18}$$

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

$$\mathbb{E}Wf(W) = \frac{1}{2\lambda}\mathbb{E}(W - W')(f(W) - f(W')) + \mathbb{E}f(W)R. \tag{1.19}$$

Proof. From (1.17) 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,

$$\begin{aligned}
\mathbb{E}W'W &= \mathbb{E}[\mathbb{E}[W'W|W]] \\
&= \mathbb{E}[W\mathbb{E}[W'|W]] \\
&= \mathbb{E}[W((1-\lambda)W + \lambda R)] \quad \text{from (1.17)} \\
&= (1-\lambda)\mathbb{E}W^2 + \lambda\mathbb{E}WR.
\end{aligned} \tag{1.20}$$

Then

$$\begin{aligned}
\mathbb{E}\Delta^2 &= \mathbb{E}(W - W')^2 \\
&= \mathbb{E}W^2 + \mathbb{E}W'^2 - 2\mathbb{E}W'W \\
&= 2\mathbb{E}W^2 - 2((1-\lambda)\mathbb{E}W^2 + \lambda\mathbb{E}WR) \quad \text{from (1.20)} \\
&= 2\lambda\mathbb{E}W^2 - 2\lambda\mathbb{E}WR.
\end{aligned} \tag{1.21}$$

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 1.5,

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

Rearranging the expression yields

$$\mathbb{E}Wf(W) = \frac{1}{2\lambda}\mathbb{E}(W - W')(f(W) - f(W')) + \mathbb{E}f(W)R. \tag{1.22}$$

□

This is just a small part of Lemma 2.4 from [4]:

Lemma 1.7. *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*

$$||f_h|| \leq 2||h'||. \tag{1.23}$$

Lemma 2.2 from [4]:

Lemma 1.8. *For fixed $z \in \mathbb{R}$ and $\Phi(z) = P(Z \leq z)$, the unique bounded solution $f_z(w)$ of the equation*

$$f'(w) - wf(w) = \mathbf{1}_{\{w \leq z\}} - \Phi(z) \tag{1.24}$$

is given by

$$f_z(w) = \begin{cases} \sqrt{2\pi} e^{w^2/2} \Phi(w)[1 - \Phi(z)] & \text{if } w \leq z \\ \sqrt{2\pi} e^{w^2/2} \Phi(z)[1 - \Phi(w)] & \text{if } w > z. \end{cases} \tag{1.25}$$

Part of Lemma 2.3 from [4]:

Lemma 1.9. *Let $z \in \mathbb{R}$ and let f_z as in (1.25) Then*

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

Generalization of Lemma 5.3 from [4]:

Lemma 1.10. *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 $a > 0$,

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

and

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

Proof. Let

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

Since

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

we have

$$\begin{aligned} 3a\lambda\sqrt{\mathbb{E}W^2} &\ge 2\lambda\mathbb{E}WF(W) \\ &= \mathbb{E}[(W - W')(f(W) - f(W'))] + 2\lambda\mathbb{E}f(W)R \quad \text{by (1.19)} \end{aligned}$$

We also bound the term involving the remainder

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

so that

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

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

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

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:

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

□

1.5 Main Theorem

This theorem will allow us to get the $\mathcal{O}N^{-1/4}$ rate under mild conditions.

Generalization of Theorem 5.5 from [4]:

Theorem 1.11. *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

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

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

$$f'(w) - wf(w) = h_{z,\alpha}(w) - \Phi(z) \tag{1.26}$$

for the smoothed indicator

$$h_{z,\alpha}(w) = \begin{cases} 1 & w \le z \\ 1 + \frac{z-w}{\alpha} & z < w \le z+\alpha \\ 0 & w > z+\alpha. \end{cases} \tag{1.27}$$

Therefore,

$$\begin{aligned} |P(W \le z) - \Phi(z)| &= |\mathbb{E}[(f'(W) - Wf(W))]| \\ &= \left|\mathbb{E}\left[f'(W) - \frac{(W'-W)(f(W') - f(W))}{2\lambda} + f(W)R\right]\right| \\ &= \left|\mathbb{E}\left[f'(W)\left(1 - \frac{(W'-W)^2}{2\lambda}\right)\right.\right. \\ &\quad \left.\left. + \frac{f'(W)(W'-W)^2 - (f(W') - f(W))(W'-W)}{2\lambda} + f(W)R\right]\right| \\ &:= |\mathbb{E}[J_1 + J_2 + J_3]| \\ &\le |\mathbb{E}J_1| + |\mathbb{E}J_2| + |\mathbb{E}J_3|. \end{aligned} \tag{1.28}$$

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

$$|\mathbb{E}J_3| \le \mathbb{E}|J_3| = \mathbb{E}|f(W)R| \le \mathbb{E}|R| \tag{1.29}$$

and

$$
\begin{aligned}
|\mathbb{E}J_1| &= \left|\mathbb{E}\left[f'(W)\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right]\right| \\
&\le \mathbb{E}\left[\left|f'(W)\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right|\right] \\
&\le \mathbb{E}\left[\left|\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right|\right] \\
&= \frac{1}{2\lambda}\mathbb{E}[|2\lambda - \mathbb{E}[(W'-W)^2|W]|] \\
&= \frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}W^2 - \mathbb{E}WR) - \mathbb{E}[(W'-W)^2|W] + 2\lambda(1-\mathbb{E}W^2+\mathbb{E}WR)|] \\
&\le \frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}W^2 - \mathbb{E}WR) - \mathbb{E}[(W'-W)^2|W]|] + \mathbb{E}|1-\mathbb{E}W^2+\mathbb{E}WR|
\end{aligned}
\tag{1.30}
$$

Note that

$$
\mathbb{E}[\mathbb{E}[(W'-W)^2|W]] = \mathbb{E}\Delta^2 = 2\lambda(\mathbb{E}W^2 - \mathbb{E}WR), \tag{1.31}
$$

so

$$
\frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}W^2 - \mathbb{E}WR) - \mathbb{E}[(W'-W)^2|W]|] \le \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])}. \tag{1.32}
$$

Combining with (1.30),

$$
\begin{aligned}
|\mathbb{E}J_1| &\le \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}|1-\mathbb{E}W^2+\mathbb{E}WR| \\
&\le \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}|1-\mathbb{E}W^2| + \mathbb{E}|WR|
\end{aligned}
\tag{1.33}
$$

Lastly, we bound the second term,

$$
\begin{aligned}
J_2 &= \frac{1}{2\lambda}(W'-W)\int_W^{W'}(f'(W)-f'(t))dt \\
&= \frac{1}{2\lambda}(W'-W)\int_W^{W'}\int_t^W f''(u)dudt \\
&= \frac{1}{2\lambda}(W'-W)\int_W^{W'}(W'-u)f''(u)du.
\end{aligned}
\tag{1.34}
$$

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,

$$\begin{aligned} |\mathbb{E}J_2| &= \left|\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)\int_W^{W'}(W'-u)f''(u)du\right]\right| \\ &= \left|\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)\int_W^{W'}\left(\frac{W+W'}{2}-u\right)f''(u)du\right]\right| \\ &\leq \left|\mathbb{E}\left[||f''||\frac{1}{2\lambda}|W'-W|\int_{\min(W,W')}^{\max(W,W')}\left|\frac{W+W'}{2}-u\right|du\right]\right| \\ &= \left|\mathbb{E}\left[||f''||\frac{1}{2\lambda}\frac{|W'-W|^3}{4}\right]\right| \\ &\leq \frac{\mathbb{E}|W'-W|^3}{4\alpha\lambda}, \end{aligned} \tag{1.35}$$

where the final inequality follows from the fact that $|h'_{z,\alpha}(x)| \leq 1/\alpha$ for all $x \in \mathbb{R}$ and Lemma 1.7.

Collecting the bounds, we obtain

$$\begin{aligned}
P(W \leq z) &\leq \mathbb{E}h_{z,\alpha}(W) \\
&\leq Nh_{z,\alpha} + \frac{\mathbb{E}|W' - W|^3}{4\alpha\lambda} + \frac{1}{2\lambda}\sqrt{\operatorname{Var}(\mathbb{E}[(W'-W)^2|W])} \\
&\quad + |1 - \mathbb{E}W^2| + \mathbb{E}|WR| + \mathbb{E}|R| \qquad (1.36) \\
&\leq \Phi(z) + \frac{\alpha}{\sqrt{2\pi}} + \frac{\mathbb{E}|W' - W|^3}{4\alpha\lambda} + \frac{1}{2\lambda}\sqrt{\operatorname{Var}(\mathbb{E}[(W'-W)^2|W])} \\
&\quad + |\mathbb{E}W^2 - 1| + \mathbb{E}|WR| + \mathbb{E}|R|
\end{aligned}$$

The minimizer of the expression is

$$\alpha = \frac{(2\pi)^{1/4}}{2}\sqrt{\frac{\mathbb{E}|W' - W|^3}{\lambda}}. \tag{1.37}$$

Plugging this in, we get the upper bound

$$\begin{aligned}
P(W \leq z) - \Phi(z) \leq (2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|W'-W|^3}{\lambda}} + \frac{1}{2\lambda}\sqrt{\operatorname{Var}(\mathbb{E}[(W'-W)^2|W])} \\
+ |\mathbb{E}W^2 - 1| + \mathbb{E}|WR| + \mathbb{E}|R|
\end{aligned} \tag{1.38}$$

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

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 [4]:

Theorem 1.12. *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

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

Proof. Now we bound $|\mathbb{E}J_2|$ with $\delta \geq 0$. From (1.28),

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

Using (1.24), $f'(W) = Wf(W) + \mathbf{1}_{\{w\leq z\}} - \Phi(z)$ and $f'(W+t) = (W+t)f(W+t) + \mathbf{1}_{\{w+t\leq z\}} - \Phi(z)$. Therefore,

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

We apply (1.9) with $w = W$, $u = 0$, and $v = t$ to get

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

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

$W' - W \leq 0$, we have

$$\begin{aligned}
\mathbb{E}J_{22}\mathbf{1}_{\{\delta \leq W'-W \leq 0\}} &= \mathbb{E}\left[(W'-W)\mathbf{1}_{\{\delta \leq W'-W \leq 0\}} \int_0^{W'-W} (\mathbf{1}_{\{W \leq z\}} - \mathbf{1}_{\{W+t \leq z\}})dt\right] \\
&= \mathbb{E}\left[(W-W')\mathbf{1}_{\{\delta \leq W'-W \leq 0\}} \int_{W'-W}^{0} (\mathbf{1}_{\{z \leq W \leq z-t\}})dt\right] \\
&\leq \mathbb{E}\left[(W-W')^2\mathbf{1}_{\{\delta \leq W'-W \leq 0\}}\mathbf{1}_{\{z-\delta \leq W \leq z\}}\right] \\
&\leq 3\delta\lambda(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|) \quad \text{by (1.10)}
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}J_{22}\mathbf{1}_{\{0 < W'-W \leq \delta\}} &= \mathbb{E}\left[(W'-W)\mathbf{1}_{\{0 < W'-W \leq \delta\}} \int_0^{W'-W} (\mathbf{1}_{\{W \leq z\}} - \mathbf{1}_{\{W+t \leq z\}})dt\right] \\
&= \mathbb{E}\left[(W'-W)\mathbf{1}_{\{0 < W'-W \leq \delta\}} \int_0^{W'-W} \mathbf{1}_{\{z-t < W \leq z\}}dt\right] \\
&\leq \mathbb{E}\left[(W'-W)^2\mathbf{1}_{\{0 < W'-W \leq \delta\}}\mathbf{1}_{\{z-\delta \leq W \leq z\}}\right] \\
&\leq 3\delta\lambda(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|) \quad \text{by (1.10)}
\end{aligned}$$

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 (1.28), noting that J_1 and J_3 stay the same. $\square$

Chapter 2

Main Proof

In this chapter, we prove the core theoretical result of this thesis: a rate of convergence bound for the randomization distribution of the t-statistic, using theorem 1.11 of chapter 1.

2.1 Motivation

Motivated by concerns regarding normality assumptions in the hypothesis being tested, Fisher [9] 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 [6] 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 [35] or David [5]. Diaconis and Lehmann [7] in their comment on Zabell's paper further expanded on some properties of these randomization tests.

Ludbrook and Dudley [22] 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 [23], and Fisher's randomization model [9]. Add some more on these two models...

Under the randomization model and using the language of triangular arrays, Lehmann [18] 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 [32], 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 [2] first obtained a Berry-Esseen bound for Student's statistic in the independent but non-identically distributed setting with additional work by Shao [31].

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

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

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,

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

$\sqrt{}$ [redacted]

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

2.3 Assumptions

Recall that the t-statistic is invariant up to sign under linear 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

$$
z_i = \sqrt{\frac{2N}{\sum(u_i - \bar{u})^2}}(u_i - \bar{u}), \tag{2.1}
$$

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:

$\sqrt{}$ [redacted]

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

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:

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

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.

In order to bound various moments of $\bar{u}_{2,\Pi}$ under the permutation distribution, we use a result of Serfling's [30]:

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

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

By assumption (2.3),

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

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

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

Therefore, by the c_r-inequality,

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

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

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

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

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

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,

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

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,

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

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

2.5 Proof

T_Π and T'_Π are exchangeable by construction:

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

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, $\pi^* = (2N, \ldots, 1) \circ \pi$. Since indices 1 to N correspond to the first group and $N+1$ to $2N$ to the second, π^* flips the groups after π, so $T_{\pi^*} = -T_\pi$.

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

For convenience, we restate theorem 1.11 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

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

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

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

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

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

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

so

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

With the preliminaries in place, we proceed to provide bounds on each term in theorem 1.11.

Proposition 2.2. $|\mathbb{E}T_\Pi^2-1|\leq c_2N^{-1}$

Proof.

$$\mathbb{E}T_\Pi^2 = \frac{N-1}{N}\mathbb{E}\left[\left(\frac{q_\Pi}{d_\Pi}\right)^2\right] \tag{2.13}$$

$$= \frac{N-1}{N}\mathbb{E}\left[\frac{4N^2\bar{u}_{2,\Pi}^2}{2N-2N\bar{u}_{2,\Pi}^2}\right] \quad \text{from (2.8)}$$

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

$$= 2(N-1)\mathbb{E}g(\bar{u}_{2,\Pi}), \tag{2.14}$$

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 (2.14) 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

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

$\square$

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

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

$$\begin{aligned} \mathrm{Var}(WV) &= \mathrm{Var}(W(V - \mathbb{E}V) + W\mathbb{E}V) \\ &\leq 2\,\mathrm{Var}(W(V - \mathbb{E}V)) + 2\,\mathrm{Var}(W\mathbb{E}V) \\ &\leq 2\mathbb{E}[W^2(V - \mathbb{E}V)^2] + 2(\mathbb{E}V)^2\,\mathrm{Var}(W) \\ &\leq 2(f_{c_2}(2))^2 N^{-2}\,\mathrm{Var}(V) + 2u_\Delta^4\,\mathrm{Var}(W). \end{aligned} \tag{2.15}$$

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

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

$$
\begin{aligned}
&\left| \mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2 \right| \\
&= \left| \mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2 + \frac{1}{(2N-1)^2} - \frac{1}{(2N-1)^2} \right| \\
&\le \left| \mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2 + \frac{1}{(2N-1)^2} \right| + \left| \frac{1}{(2N-1)^2} \right| \\
&\le \left| \mathbb{E}\left[\left(\frac{1}{1-\bar{u}_{2,\Pi}^2}\right)^2\right] - 1 - \frac{2}{2N-1} - \left(\left(\mathbb{E}\left[\frac{1}{1-\bar{u}_{2,\Pi}^2}\right]\right)^2 - \left(1+\frac{1}{2N-1}\right)^2\right) \right| + \left| \frac{1}{(2N-1)^2} \right| \\
&\le c_4 N^{-2} + 2\left(1+\frac{1}{2N-1}\right) c_5 N^{-2} + c_5^2 N^{-4} + \left| \frac{1}{(2N-1)^2} \right| \\
&\le (c_4 + 3c_5 + c_5^2 + \frac{1}{4}) N^{-2} \\
&:= c_6 N^{-2}
\end{aligned}
$$

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

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

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

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

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

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

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

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,

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

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

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

Applying to the second term,

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}\left(\mathbb{E}\left[\left(T'_\Pi \frac{d_\Pi - d'_\Pi}{d_\Pi}\right)^2 \middle| \Pi = \pi\right]\right) \\
&\le \mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}(d_\Pi - d'_\Pi)\right)^4\right]
\end{aligned}$$

$\sqrt{}$ [redacted]

$$\begin{aligned}
&\le \quad \mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^8\right] \mathbb{E}[(d_\Pi - d'_\Pi)^8] \\
&\le \sqrt{f_{c_6}(8) N^{-8/2} f_{c_4}(8) N^{-8}} \text{ from (2.10), (2.7)} \\
&= \sqrt{f_{c_6}(8) f_{c_4}(8)} N^{-6} \\
&:= c_7 N^{-6}
\end{aligned}$$

And to the third,

$$\begin{aligned}
\mathrm{Var}(Z) &= 4\,\mathrm{Var}\left(\mathbb{E}\left[\left(\frac{2u_{\Pi(J)} - 2u_{\Pi(I)}}{d_\Pi} T'_\Pi \frac{d_\Pi - d'_\Pi}{d_\Pi}\right)\middle|\Pi = \pi\right]\right) \\
&\leq 16u_\Delta^2 \mathbb{E}\left[\left(\frac{1}{d_\Pi}\frac{q'_\Pi}{d_\Pi d'_\Pi}(d_\Pi - d'_\Pi)\right)^2\right]
\end{aligned}$$

[redacted]

$$\begin{aligned}
&\leq 16u_\Delta^2 f_{c_2}(2)N^{-2/2} \quad \mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^4\right]\mathbb{E}[(d_\Pi - d'_\Pi)^4] \text{ from (2.5)} \\
&\leq 16u_\Delta^2 f_{c_2}(2)N^{-1}\sqrt{f_{c_6}(4)N^{-4/2}f_{c_4}(4)N^{-4}} \text{ from (2.10), (2.7)} \\
&\leq 16u_\Delta^2 f_{c_2}(2)(f_{c_6}(4))^{-1/2}(f_{c_4}(4))^{-1/2}N^{-4} \\
&:= c_8 N^{-4}
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(T'_\Pi - T_\Pi)^2|T_\Pi])} \\
&= N\sqrt{(\mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z))}
\end{aligned}$$

[redacted]

$$\begin{aligned}
&\leq N \quad 8(f_{c_2}(2))^2\left(20 + 16\left(\sum_{i=1}^{2N} u_i^4\right)N^{-2}\right)N^{-4} + 8u_\Delta^4\frac{c_6}{4}N^{-4} + c_7N^{-6} + c_8N^{-4} \\
&:= N^{-1}c_3\sqrt{20 + 16\frac{\sum_{i=1}^{2N} u_i^4}{N^2}}
\end{aligned}$$

□

Proposition 2.4. $(2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|T'_\Pi - T_\Pi|^3}{\lambda}} < (2\pi)^{-1/4}c_9 N^{-1/4}$.

Proof. The strategy is to break apart the remainder term from the main piece. From

(2.11),

$$\mathbb{E}|T'_\Pi - T_\Pi|^3 = \left(\frac{N-1}{N}\right)^{3/2} \mathbb{E}\left[d_\Pi^{-3}\left|2u_{\Pi(J)} - 2u_{\Pi(I)} + q'_\Pi \frac{d_\Pi - d'_\Pi}{d'_\Pi}\right|^3\right]$$

$$\leq 8\left(8u_\Delta^3 \mathbb{E}[d_\Pi^{-3}] + \sqrt{\text{[redacted]}\ \mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^6\right]\mathbb{E}[(d_\Pi - d'_\Pi)^6]}\right)$$

$$\leq 64u_\Delta^3 f_{c_2}(3)N^{-3/2} + 8\sqrt{f_{c_6}(6)N^{-6/2}f_{c_4}(6)N^{-6}} \text{ from } (2.5), (2.10), (2.7)$$

$$\leq \frac{c_9^2}{2}N^{-3/2}$$

Therefore,

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

□

Proposition 2.5. $\mathbb{E}|R| \leq \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}$.

Proof.

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

$$\leq \frac{N}{2}\mathbb{E}\left|\frac{q'_\Pi}{d_\Pi d'_\Pi}(d_\Pi - d'_\Pi)\right|$$

$$\leq \frac{N}{2}\sqrt{\mathbb{E}\left|\frac{q'_\Pi}{d_\Pi d'_\Pi}\right|^2 \mathbb{E}[d_\Pi - d'_\Pi]^2}$$

$$\leq \frac{N}{2}\sqrt{f_{c_6}(2)N^{-2/2}f_{c_4}(2)N^{-2}} \text{ from } (2.10), (2.7)$$

$$= \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}$$

$\square$

Proposition 2.6. $\mathbb{E}|T_\Pi R| \leq \frac{1}{2}(f_{c_6}(4) f_{c_4}(4))^{1/4}\sqrt{2+2c_1}N^{-1/2}$.

Proof.

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

$$\leq \frac{N}{2}\mathbb{E}\left|T_\Pi \frac{q'_\Pi}{d_\Pi d'_\Pi}(d_\Pi - d'_\Pi)\right|$$

[redacted]

$$\leq \frac{N}{2} \quad \mathbb{E}T_\Pi^2 \mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^2 (d_\Pi - d'_\Pi)^2\right]$$

[redacted]

$$\leq \frac{N}{2} \quad \mathbb{E}T_\Pi^2 \quad \mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^4\right]\mathbb{E}[(d_\Pi - d'_\Pi)^4]$$

$$\leq \frac{N}{2}\sqrt{\mathbb{E}T_\Pi^2 \sqrt{f_{c_6}(4)N^{-4/2} f_{c_4}(4)N^{-4}}} \text{ from (2.10), (2.7)}$$

$$= \frac{N^{-1/2}}{2}(f_{c_6}(4) f_{c_4}(4))^{1/4}\sqrt{\mathbb{E}T_\Pi^2}$$

$$\leq \frac{1}{2}(f_{c_6}(4) f_{c_4}(4))^{1/4}\sqrt{2+2c_1}N^{-1/2}$$

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

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

$$\begin{aligned}\sup_{t\in\mathbb{R}}|P(T_\Pi\le t)-\Phi(t)| &\le (2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|T'_\Pi-T_\Pi|^3}{\lambda}}+\frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(T'_\Pi-T_\Pi)^2|T_\Pi])}\\ &\quad+|\mathbb{E}T_\Pi^2-1|+\mathbb{E}|T_\Pi R_\Pi|+\mathbb{E}|R_\Pi|\\ &\le (2\pi)^{-1/4}c_9N^{-1/4}+N^{-1}c_3\sqrt{20+16\frac{\sum_{i=1}^{2N}u_i^4}{N^2}}+c_2N^{-1}\\ &\quad+\frac{1}{2}(f_{c_6}(4)f_{c_4}(4))^{1/4}\sqrt{2+2c_1}N^{-1/2}+\frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}\end{aligned}$$

2.6 Better Rate

Here, we use theorem 1.12 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 2.2, $\mathbb{E}T_\Pi^2\le c_2N^{-1}+1$, and from proposition 2.5, $\mathbb{E}|R|\le \frac{1}{2}\sqrt{f_{c_6}(2)f_{c_4}(2)}N^{-1/2}$. If $\delta<c_{10}N^{-1/2}$ for N sufficiently large, applying theorem 1.12, we see

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

It still remains to be seen 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 $u_i = i$,

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

Chapter 3

Simulations

This chapter is a computational companion to chapter 2.

3.1 Preliminaries

First, we provide simulations accompanying section 2.4. We generate i.i.d. samples $\{u_i\}_{i=1}^N \sim \mathcal{N}(-1,1)$ and $\{u_i\}_{i=N+1}^{2N} \sim \mathcal{N}(1,1)$ for exponentially-spaced values of $N \in \{\text{floor}(10^{.5+.5i})\}_{i=1}^7$. The u_i are scaled and centered, and for each N, we perform $10,000$ permutations.

We plot Monte Carlo estimates of the means of each term, scaled by the rate of our bound, along with 95%ile bootstrap confidence intervals for different values of $p \in \{2,4,6,8\}$.

Due to the flatness of the curves, we conclude that the bounds we have proved are of the correct rate. In addition, we can observe the behavior of the constants as functions of p. For instance, our $f_{c_3}(p)$ constant for $\mathbb{E}h_{\Pi}^p$ appears to be an exponential function of p.

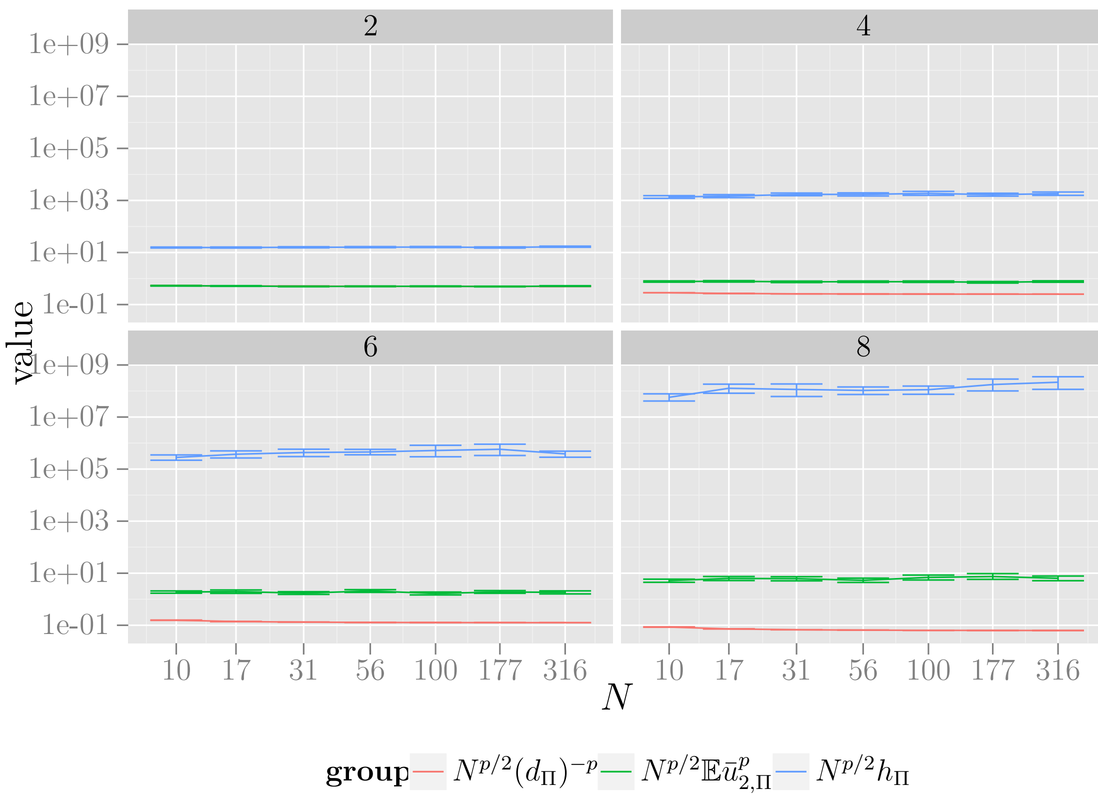


Figure 3.1: Log-log plots of values scaled by proven upper bounds of rates, faceted on p.

Here, to compute the corresponding "prime" random variables in the coupled pair, in each permutation we pick a transposition uniformly at random among transpositions that switch groups.

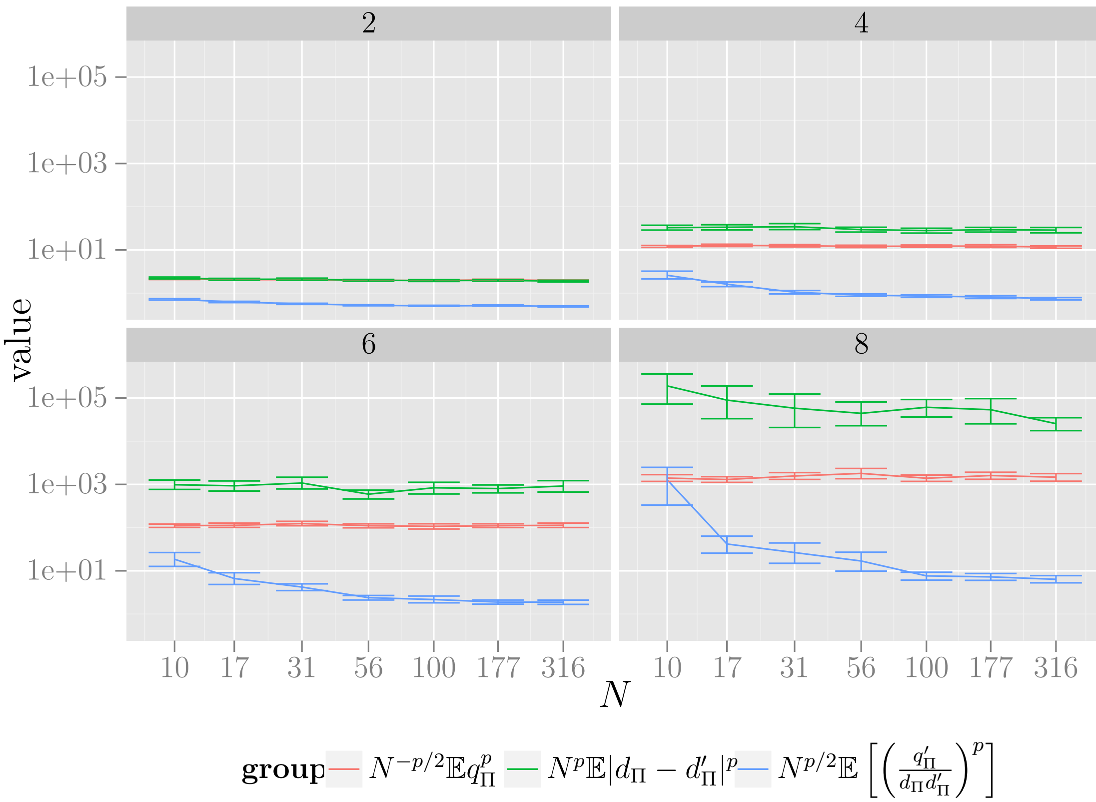


Figure 3.2: Log-log plots of values scaled by proven upper bounds of rates, faceted on p.

It is possible that the bound of rate $N^{-p/2}$ on $\mathbb{E}\left[\left(\frac{q'_\Pi}{d_\Pi d'_\Pi}\right)^p\right]$ is a bit conservative.

3.2 Approximate Regression Condition

From the approximate regression condition

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

we get

$$\mathbb{E}[T'_\Pi | T_\Pi] = (1-\lambda)T_\Pi - \lambda R_\Pi.$$

That is, the conditional expectation of T'_Π on T_Π is expected to lie near the line $(1-\lambda)T_\Pi$ with a small perturbation of order $1/N$ (recall that $\lambda = 2/N$).

For various values of N, we compute 20 permutations that correspond with 20 values of T_Π. For each T_Π, we draw a transposition (I, J) uniformly at random from the space of our allowable transpositions, repeating this 100 times.

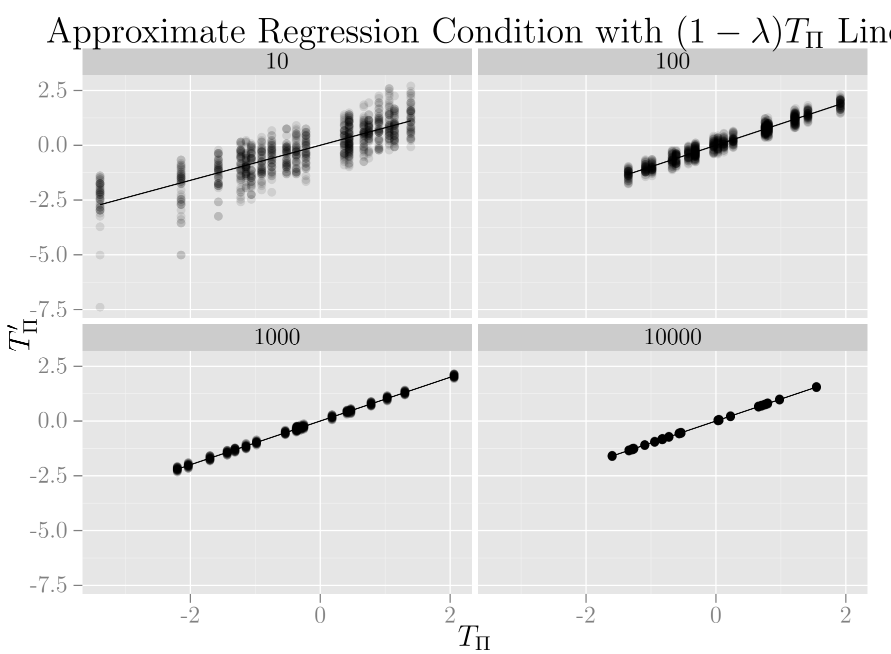


Figure 3.3: Faceted on per-group sample size, N.

The approximate regression condition appears to hold visually.

3.3 Main Bounds

Here we simulate the main bounds under the same setting as the previous section.

3.3.1 Failure of Monte Carlo

Again, we simulate the conditional expectations of the form $\mathbb{E}[f(T'_\Pi, T_\Pi)|T_\Pi]$ with 1,000 draws from the uniform distribution on all group-switching transpositions (I, J).

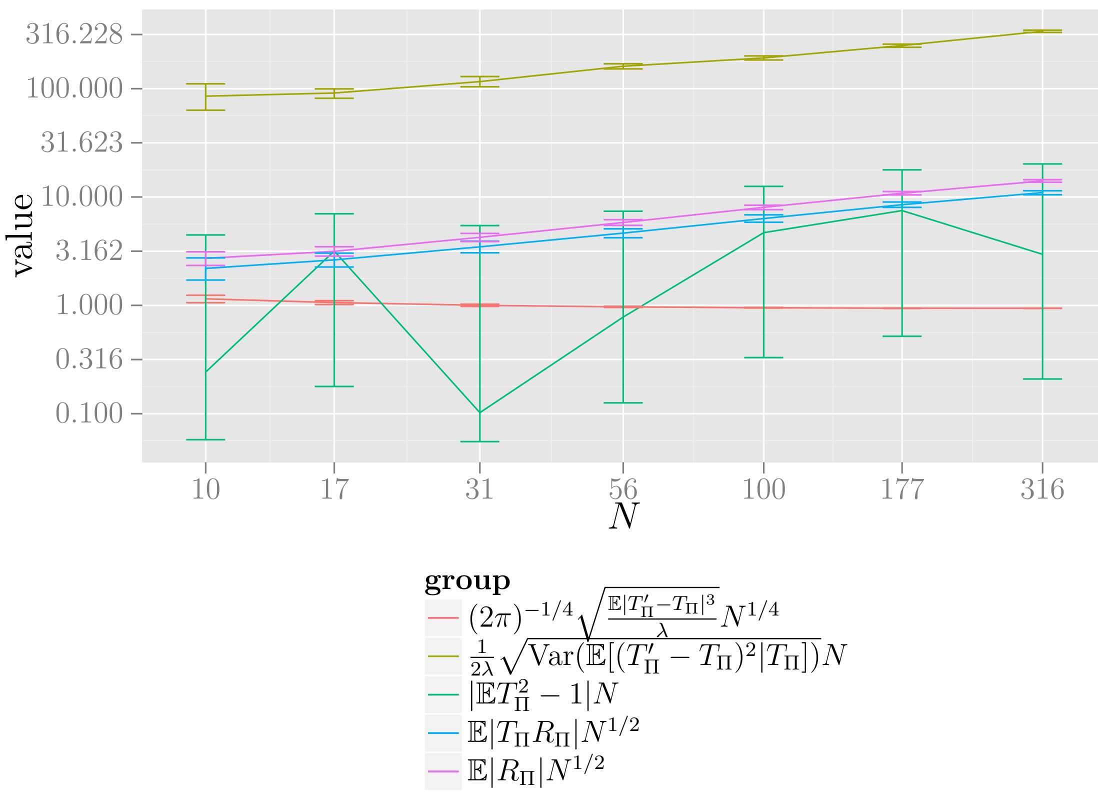


Figure 3.4: Log-log plot of values for each term in the bound, simulating the conditional by Monte Carlo.

The MC error is too large, and we see some scaled bounds actually increase.

3.3.2 Exact Conditional Expectation Calculations

Here we plot T' for all N^2 group-switching transpositions (I, J):

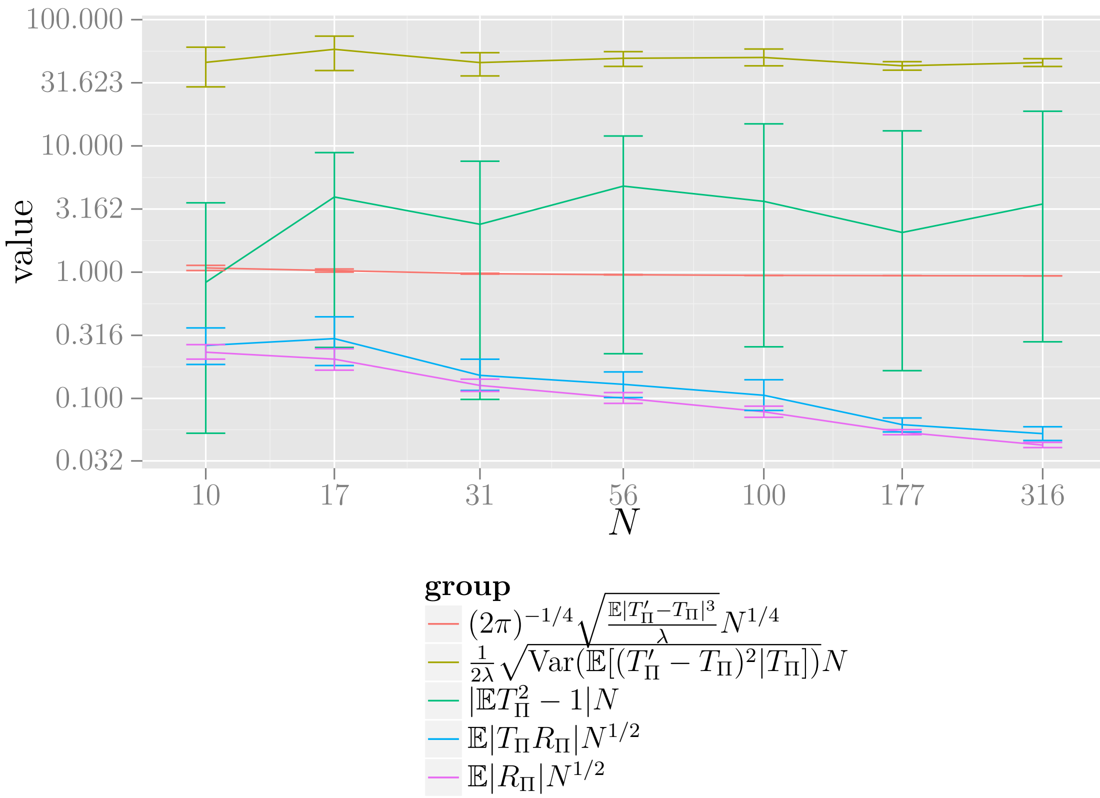


Figure 3.5: Log-log plot of values for each term in the bound, calculating the conditional expectation exactly ($10N$ permutations each).

Our bounds appear to be of the correct order or slightly conservative in some cases. The bounds on the remainder terms ($\mathbb{E}|R_\Pi|$ and $\mathbb{E}|T_\Pi R_\Pi|$) are of order $N^{1/2}$, but the true rates are probably lower.

3.3.3 Better Rate

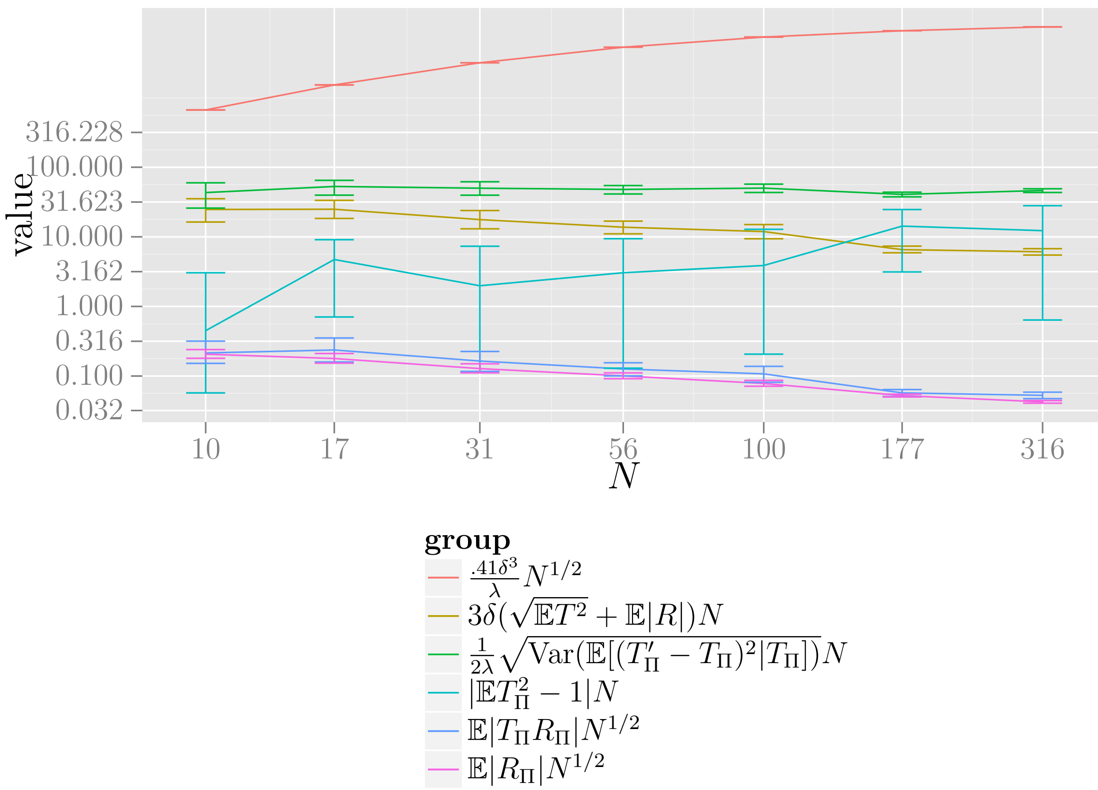
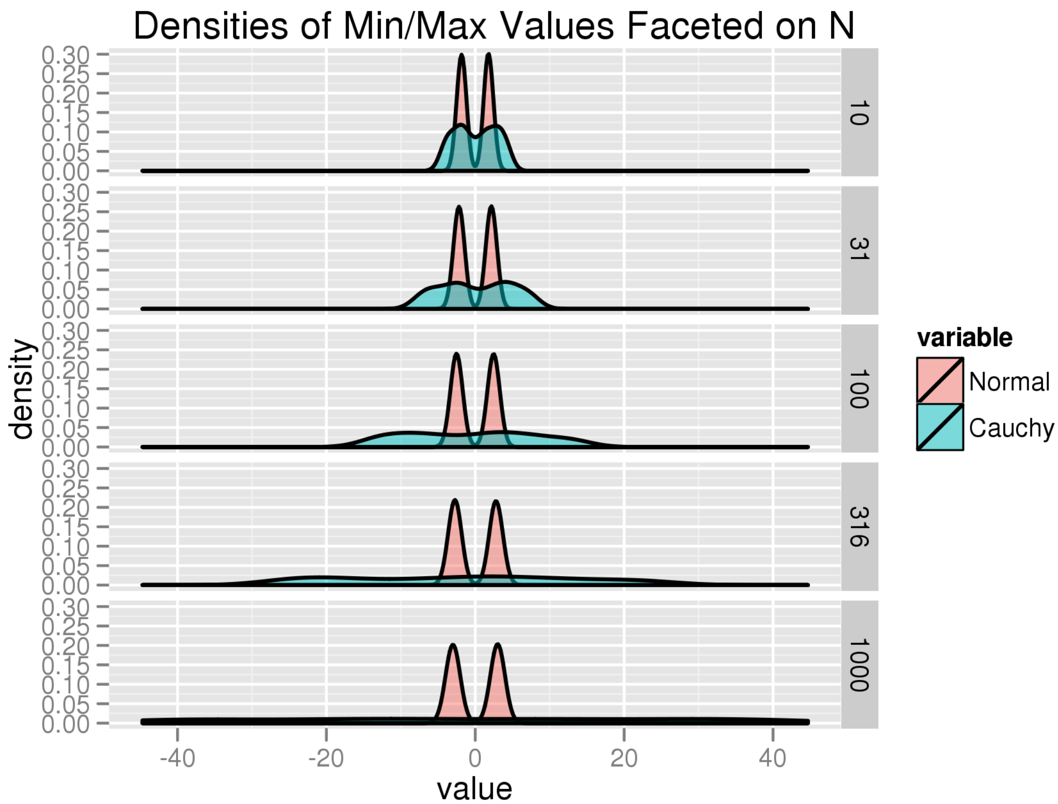


Figure 3.6: Log-log plot of values for each term in the bound, calculating the conditional expectation exactly ($10N$ permutations each).

3.4 True Rate

To assess the true rate of convergence, we consider two settings: the earlier Normal setting and group draws from a Cauchy distribution with location parameters -1 and 1 depending on the group. Our bounds include a dependence on $u_\Delta = \max_i u_i - \min_i u_i$. To better understand the differences between these two models, we simulate the minimum and maximum scaled (mean 0 and sum of squares $2N$) values:



For $N = 1000$, the Normal model typically has u_Δ values around 6. In contrast, the Cauchy model has u_Δ values closer to 40.

Here, we plot the empirical Kolmogorov-Smirnov test statistic in the following three settings:

1. a standard Normal draw of size N (repeated N times to get the empirical distribution)

2. the permutation t-statistic under Cauchy sampling (N permutations)

3. the permutation t-statistic under Normal sampling (N permutations)

We also add the sum of the five unscaled, simulated bound terms ($200,000$ permutations) from the previous section.

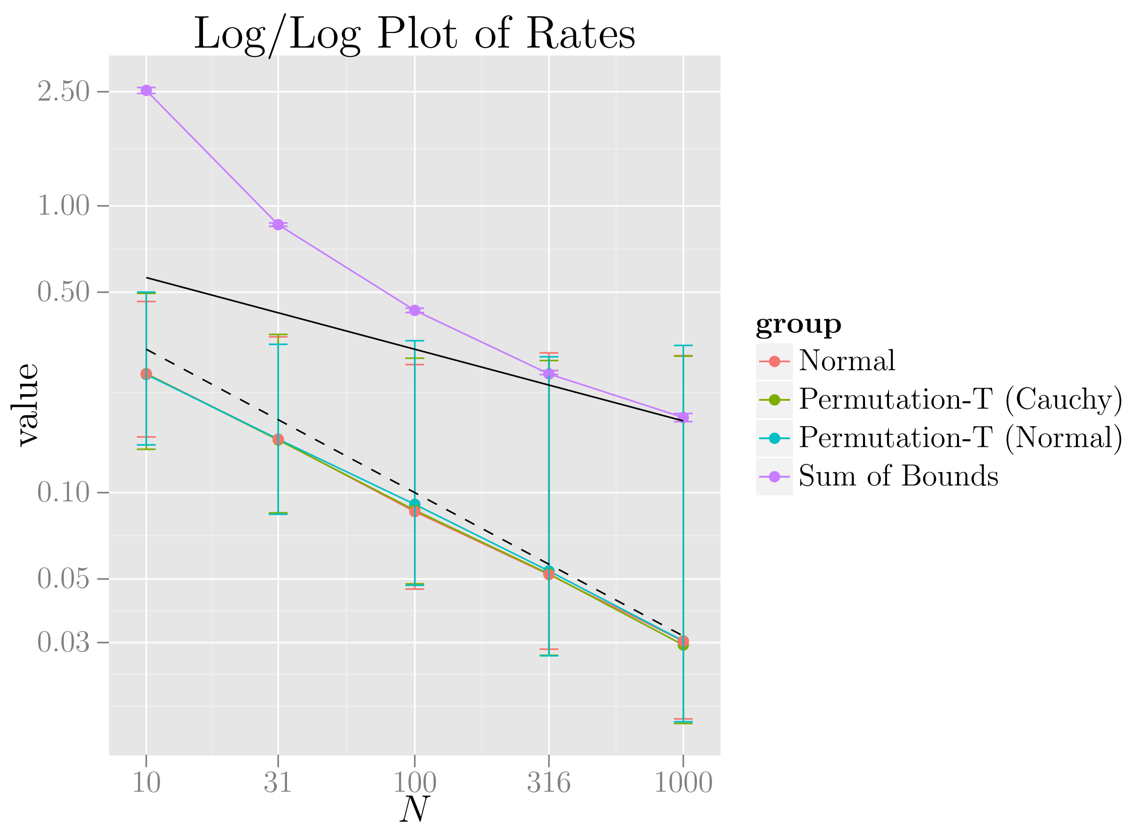


Figure 3.7: Solid black line: $N^{-1/4}$; dashed black line: $N^{-1/2}$

It's not a fair comparison to place the sum of the bounds on the same plot because that was computed over $200,000$ separate permutations instead of the 500 shared by the other three. Still, we can draw some general conclusions. The normal and two permutation-t K-S statistics decay perfectly at a rate of $N^{-1/2}$, and our bound follows a rate of $N^{-1/4}$, suggesting that the true rate of convergence is the former. Also, the error-bars seem to be increasing in size but are actually roughly constant due to the log-log scale.

Chen et al. [4] provide a simple example (pp.154-155) in which the sum of i.i.d. random variables yields

$$E|W' - W|^3 = \frac{4}{N^{3/2}}$$

with $\lambda = N^{-1}$. This leads to an $O(N^{-1/4})$ bound, which is suboptimal and apparently not uncommon when applying this kind of theorem.

3.5 Efficient Updates

Instead of conditioning on the value of T_Π, we condition on the observed permutation π. For N observations in each group, there are $N^2 T'_\Pi$ values that come from swapping one value in the first group with one value in the second. T'_Π should not differ much from T_Π, and calculating the t-statistics from scratch is inefficient.

We use an efficient t-statistic update rule to easily calculate millions of t-statistics. The two sample t-statistic is given by

$$T_\Pi = \frac{\bar{x} - \bar{u}}{\sqrt{\frac{2}{n}}\sqrt{\frac{1}{2}(S_X^2 + S_U^2)}},$$

where $S_X^2 = \frac{1}{N-1}(\sum_{i=1}^N x_i^2 - n\bar{x}^2)$.

Let T_{x_i,u_j} be the result of T' by swapping x_i with u_j:

$$
\begin{aligned}
\Delta &\equiv u_j - x_i \\
\bar{x}_{x_i,u_j} &= \bar{x} - \frac{1}{N}x_i + \frac{1}{N}u_j = \bar{x} + \frac{\Delta}{N} \\
\bar{u}_{x_i,u_j} &= \bar{u} + \frac{1}{N}x_i - \frac{1}{N}u_j = \bar{u} - \frac{\Delta}{N} \\
S^2_{X_{x_i,u_j}} &= \frac{1}{N-1}(\sum_{k=1}^{N} x_k^2 - x_i^2 + u_j^2) - \frac{N}{N-1}\bar{x}^2_{x_i,u_j} \\
S^2_{U_{x_i,u_j}} &= \frac{1}{N-1}(\sum_{k=1}^{N} u_k^2 + x_i^2 - u_j^2) - \frac{N}{N-1}\bar{u}^2_{x_i,u_j} \\
\bar{x}^2_{x_i,u_j} &= \bar{x}^2 + \frac{2\Delta}{N}\bar{x} + \frac{\Delta^2}{N} \\
\bar{u}^2_{x_i,u_j} &= \bar{u}^2 - \frac{2\Delta}{N}\bar{u} + \frac{\Delta^2}{N}
\end{aligned}
$$

Then

$$
\begin{aligned}
T_{x_i,u_j} &= \frac{\bar{x}_{x_i,u_j} - \bar{u}_{x_i,u_j}}{\sqrt{\frac{2}{N}}\sqrt{\frac{1}{2}(S^2_{X_{x_i,u_j}} + S^2_{U_{x_i,u_j}})}} \\
&= \frac{\bar{x} - \bar{u} + \frac{2\Delta}{N}}{\sqrt{\frac{2}{N}}\sqrt{\frac{1}{2(N-1)}[\sum_{k=1}^{N}(x_k^2 + u_k^2) - N(\bar{x}^2 + \bar{u}^2 + \Delta(\frac{2\bar{x}}{n} - \frac{2\bar{u}}{n}) + \frac{2}{n^2}\Delta^2)]}}.
\end{aligned}
$$

Only the terms involving Δ need to be recomputed for each of the N^2 swaps.

Consider a naïve implementation based on a double for-loop and recomputing each t-statistic anew versus a vectorized approach using the update formula:

```
computeAllCond2 <- function(T, N, u, l, x, y){
    minus <- which(l == -1)
    plus <- which(l == 1)
    Tprime <- 1:(N^2)
    for(j in 1:N){
      for(k in 1:N){
        swap <- c(minus[j], plus[k])
```

```
        l[swap] <- l[rev(swap)]
        Tprime[N*(j-1)+k] <- t.test(u[l==1], u[l==-1], var.equal=TRUE)$statistic
        l[swap] <- l[rev(swap)]
      }
    }
    data.frame("T" = T, "Tprime" = Tprime, "N" = N, "lambda" = 2 / N)
}

computeAllCond <- function(T, N, u, l, x, y){
  del <- rep(y, length(x)) - rep(x, each = length(y))
  xbar <- mean(x)
  ybar <- mean(y)
  Tprime <- -(xbar - ybar + 2/N*del) /
    (sqrt(2/N)*sqrt(sum(u^2)/(2*(N-1)) - 1/2*N/(N-1)*(xbar^2 + ybar^2 + 2*del/N*(xbar-yb
  data.frame("T" = T, "Tprime" = Tprime, "N" = N, "lambda" = 2 / N)
}
```

We observe roughly a 2,000 times increase in speed on a problem instance of size $N = 100$. With byte-compilation and additional tuning, a four order of magnitude increase is possible.

```
> system.time(computeAllCond2(T, N, u, l, x, y))
   user  system elapsed
  7.333   0.000   7.334
> system.time(computeAllCond(T, N, u, l, x, y))
   user  system elapsed
  0.005   0.000   0.004
> sum((sort(computeAllCond(T, N, u, l, x, y)$Tprime) - sort(computeAllCond2(T, N, u, l,
[1] 3.137579e-27
dat <- ldply(rep(floor(10^(seq(1, 3.5, by=.5))), each = 8),
simulateBounds, .parallel = TRUE, .progress = "text")
> print(object.size(dat), units = "Gb")
```

```
2.6 Gb
```

3.6 A Different Exchangeable Pair

Rather than only consider transpositions that swap one element of the first group with one from the second group, we have a few different choices. Let's take the other extreme, where we consider all $(2N)^2$ transpositions, including null transpositions. There are N^2 transpositions within each group, for a total of $2N^2$. Each of these does not change the t-statistic. We previously only considered the N^2 transpositions where $I < J$. There are another N^2 with $I > J$. These transpositions have exactly the same effect as the previous group $(I, J) = (J, I)$, and all within-group transpositions have no effect.

The only changes should be to adjust the weights when taking conditional expectations (the weights should be $1/2$) and to divide λ by 2. The new λ is N^{-1}.

However, every term involving the conditional expectation also has a division by lambda, so any decrease in the c.e. is cancelled out by a corresponding decrease in λ, so there is no change in any of the simulations.

It's nice that the calculations are invariant to change in the exchangeable pair. Whether that holds true for more drastic changes (e.g. swapping more than 2 elements) is not known.

3.7 Better Rate

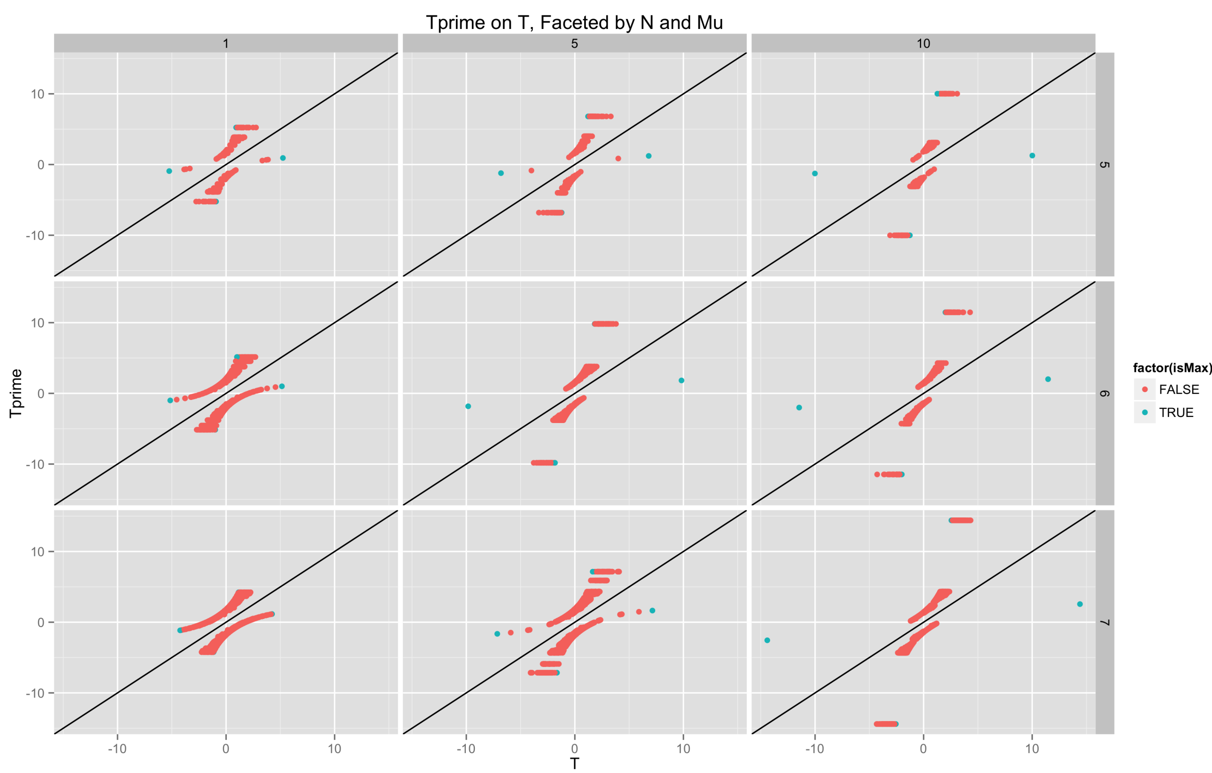
We observe two samples with equal sample size: $S_1 = \{u_i\}_{i=1}^N$ and $S_2 = \{u_i\}_{i=N+1}^{2N}$. Student's two-sample t-statistic is given by

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

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. And there are probably some tricks we can apply to reduce the N^2. But this still doesn't seem to be very tractable.

3.8 T and T'

Let's first plot all possible values of T, and the corresponding value of T' that maximizes $|T - T'|$. Here, we make N draws of sample 1 from $\mathcal{N}(\mu, 1)$ and N draws of sample 2 from $\mathcal{N}(0, 1)$, varying $N \in \{5, 6, 7\}$ and $\mu \in \{1, 2, 5\}$, coloring the (T, T') pair that maximizes $|T - T'|$. There is some symmetry (the pair shows up 4 times) because of swapping T with $-T$ (2 swaps) and T' with T (2 swaps).



Unfortunately, we can't make N much bigger than 7 using the current technique.

3.9 Shortcut

It always seems to be the case that, say, the minimum (equivalently, the maximum due to symmetry) value of T_π maximizes $|T - T'|$. Knowing the permutation π that maximizes $|T - T'|$, we can try to figure out the corresponding transposition (i, j).

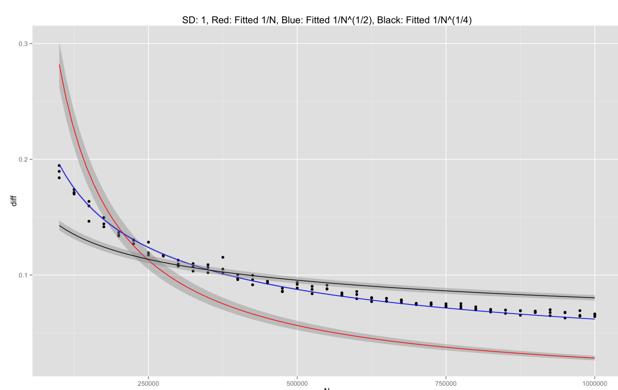
It seems reasonable that sorting the data into its order statistics $\{u_{(i)}\}_{i=1}^{2N}$ will minimize T. That is, let the N smallest be in the first group and the next N in the second.

Another thing that seems reasonable to find the (i, j) that maximizes $|T - T'|$ is to swap the "most different" sample of the first group with that of the second group: $u_{(1)}$ with $u_{(2N)}$.

I tried this shortcut (it only really involves sorting the data) and compared it with the exact methodology of the above section and found agreement in all the tested settings. This lets us really ramp up N in our simulations.

I haven't really tried to prove it yet: it looks challenging.

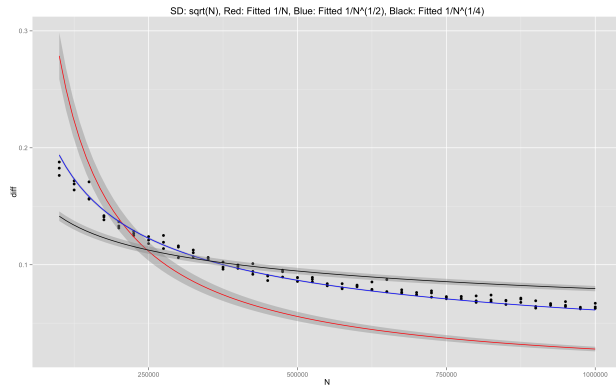
Consider drawing sample 1 from $\mathcal{N}(2, \sigma^2)$ and sample 2 from $\mathcal{N}(0, \sigma^2)$ where $\sigma = 1$:

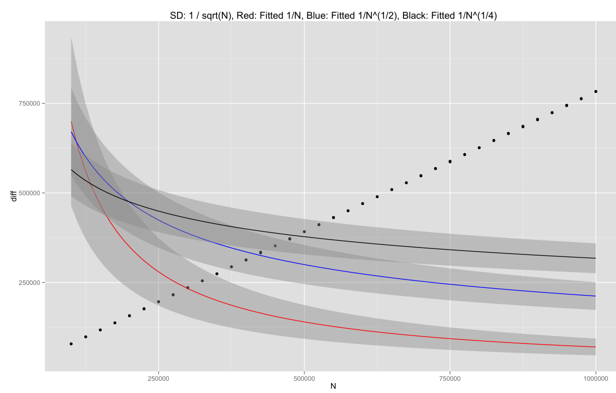


Now consider $\sigma = \sqrt{N}$ (so the power is constant in the sample size):

Finally, $\sigma = \frac{1}{\sqrt{N}}$:

The last situation is one of those pathological cases that we were trying to avoid, where the within-group variance vanishes so the distributions tend toward point masses. But the other reasonable cases look good. That is, if the shortcut works, it





appears that $\delta = \max_{\pi,i,j} |T_\pi - T_{\pi\circ(i,j)}|$ is $O(N^{-1/2})$ for these reasonable cases.

References

[1] A. Anonymous and B. Anonymous. A rate of convergence bound for the randomization t-distribution using stein's method. *Unpublished Technical Report.*

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

[3] K.M. Borgwardt, A. Gretton, M.J. Rasch, H.P. Kriegel, B. Schölkopf, and A.J. Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006.

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

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

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

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

[8] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007.

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

[10] J.H. Friedman. On Multivariate Goodness–of–Fit and Two–Sample Testing. *Proceedings of Phystat2003, http://www.slac.stanford.edu/econf/C*, 30908, 2003.

[11] Jeff Gentry. *twitteR: R based Twitter client*, 2011. R package version 0.99.6.

[12] M. Gönen and E. Alpaydın. Multiple kernel learning algorithms. *Journal of Machine Learning Research*, 12:2211–2268, 2011.

[13] A. Gretton, KM Borgwardt, M. Rasch, B. Schölkopf, and AJ Smola. A kernel method for the two-sample-problem. *Advances in Neural Information Processing Systems*, 19:513–520, 2007.

[14] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13:723–773, 2012.

[15] A. Gretton, K. Fukumizu, Z. Harchaoui, and B.K. Sriperumbudur. A fast, consistent kernel two-sample test. *Advances in Neural Information Processing Systems*, 22:673–681, 2010.

[16] Alexandros Karatzoglou, Alex Smola, Kurt Hornik, and Achim Zeileis. kernlab – an S4 package for kernel methods in R. *Journal of Statistical Software*, 11(9):1–20, 2004.

[17] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L.E. Ghaoui, and M.I. Jordan. Learning the kernel matrix with semidefinite programming. *The Journal of Machine Learning Research*, 5:27–72, 2004.

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

[19] E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses*. Springer Verlag, 2005.

[20] C. Leslie, E. Eskin, and W.S. Noble. The spectrum kernel: A string kernel for SVM protein classification. In *Proceedings of the Pacific Symposium on Biocomputing*, volume 7, pages 566–575. Hawaii, USA., 2002.

[21] H.T. Lin, C.J. Lin, and R.C. Weng. A note on platts probabilistic outputs for support vector machines. *Machine learning*, 68(3):267–276, 2007.

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

[23] 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.

[24] M. Panov, A. Tatarchuk, V. Mottl, and D. Windridge. A modified neutral point method for kernel-based fusion of pattern-recognition modalities with incomplete data sets. *Multiple Classifier Systems*, pages 126–136, 2011.

[25] J. Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

[26] N. Poh, D. Windridge, V. Mottl, A. Tatarchuk, and A. Eliseyev. Addressing missing values in kernel-based multimodal biometric fusion using neutral point substitution. *Information Forensics and Security, IEEE Transactions on*, 5(3):461–469, 2010.

[27] S. Qiu and T. Lane. Multiple kernel learning for support vector regression. *Computer Science Department, The University of New Mexico, Albuquerque, NM, USA, Tech. Rep*, 2005.

[28] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.

[29] B. Schölkopf and A.J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. the MIT Press, 2002.

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

[31] Q.M. Shao. An explicit berry-esseen bound for students t-statistic via steins method. *Steins Method and Applications (AD Barbour and LHY Chen eds). Lecture Notes Series, Institute for Mathematical Sciences, NUS*, 5:143–155, 2005.

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

[33] G. Wahba. Multivariate function and operator estimation, based on smoothing splines and reproducing kernels. In *SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-*, volume 12, pages 95–95. ADDISON-WESLEY PUBLISHING CO, 1992.

[34] G. Wahba et al. Support vector machines, reproducing kernel hilbert spaces and the randomized gacv. *Advances in Kernel Methods-Support Vector Learning*, 6:69–87, 1999.

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