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
201?

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.

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 proof of the Hoeffding combinatorial central limit theorem (HC-CLT) 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.

1.2 Hoeffding combinatorial CLT

Theorem 1.1. *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 $n-1$:*

$$\sum_{j=1}^{n} a_{ij} = 0 \tag{1.1}$$

$$\sum_{i=1}^{n} a_{ij} = 0 \tag{1.2}$$

$$\sum_{i=1,j=1}^{n} a_{ij}^2 = n-1 \tag{1.3}$$

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

$$P(\Pi = \pi) = \frac{1}{n!}. \tag{1.4}$$

Define

$$W = \sum_{i=1}^{n} a_{i\Pi(i)} \tag{1.5}$$

to be the sum of a random diagonal. Then

$$|P(W \le w) - \Phi(w)| \le \frac{C}{\sqrt{n}} \left[\sqrt{\sum_{i,j=1}^{n} a_{ij}^4} + \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^3} \right]. \tag{1.6}$$

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:

$$P(I = i, J = j) = \frac{1}{n(n-1)} \quad i, j \in \{1, \dots, n\}, i \ne j. \tag{1.7}$$

Define the random permutation Π' by

$$\Pi'(i) = \Pi \circ (I, J) = \begin{cases} \Pi(J) & i = I \\ \Pi(I) & i = J \\ \Pi(i) & \text{else.} \end{cases} \tag{1.8}$$

We construct our exchangeable pair by defining

$$W' = \sum_{i=1}^{n} a_{i\Pi'(i)} = W - a_{I\Pi(I)} + a_{I\Pi(J)} - a_{J\Pi(J)} + a_{J\Pi(I)}. \tag{1.9}$$

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

$$\lambda = \frac{2}{n-1}. \tag{1.10}$$

To bound the variance component, compute

$$
\begin{aligned}
\mathbb{E}[(W-W')^2|\Pi] &= \mathbb{E}[(a_{I\Pi(I)} - a_{I\Pi(J)} + a_{J\Pi(J)} - a_{J\Pi(I)})^2|\Pi] \\
&= \mathbb{E}[a_{I\Pi(I)}^2 + a_{J\Pi(J)}^2 + a_{I\Pi(J)}^2 + a_{J\Pi(I)}^2 \\
&\quad - 2a_{I\Pi(I)}a_{I\Pi(J)} - 2a_{J\Pi(J)}a_{J\Pi(I)} - 2a_{I\Pi(I)}a_{J\Pi(I)} - 2a_{J\Pi(J)}a_{I\Pi(J)} \\
&\quad + 2a_{I\Pi(I)}a_{J\Pi(J)} + 2a_{I\Pi(J)}a_{J\Pi(I)}|\Pi] \\
&= \frac{2}{n}\sum_{i=1}^{n} a_{i\Pi(i)}^2 + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(j)}^2 \\
&\quad - \frac{4}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(i)}a_{i\Pi(j)} - \frac{4}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(i)}a_{j\Pi(i)} \\
&\quad + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(i)}a_{j\Pi(j)} + \frac{2}{n}\frac{1}{n-1}\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(j)}a_{j\Pi(i)} \\
&= \frac{2}{n}\sum_{i=1}^{n} a_{i\Pi(i)}^2 + \frac{2}{n}\frac{1}{n-1}\left(\sum_{i,j=1}^{n} a_{i\Pi(j)}^2 - \sum_{i=1}^{n} a_{i\Pi(i)}^2\right) \\
&\quad - \frac{4}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(a_{i\Pi(i)}\sum_{j=1}^{n}\left(a_{i\Pi(j)} + a_{j\Pi(i)}\right) - 2a_{i\Pi(i)}^2\right) \\
&\quad + \frac{2}{n}\frac{1}{n-1}\left(\sum_{i,j=1,i\neq j}^{n} a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right) \\
&= \frac{2}{n}\left(1 - \frac{1}{n-1}\right)\sum_{i=1}^{n} a_{i\Pi(i)}^2 + \frac{2}{n} \\
&\quad + \frac{8}{n}\frac{1}{n-1}\sum_{i=1}^{n} a_{i\Pi(i)}^2 \\
&\quad + \frac{2}{n}\frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)}\right) - \frac{4}{n}\frac{1}{n-1}\sum_{i=1}^{n} a_{i\Pi(i)}^2 \\
&= \frac{2}{n} + \frac{2(n+2)}{n(n-1)}\sum_{i=1}^{n} a_{i\Pi(i)}^2 + \frac{2}{n(n-1)}\sum_{i,j=1,i\neq j}^{n}(a_{i\Pi(i)}a_{j\Pi(j)} + a_{i\Pi(j)}a_{j\Pi(i)})
\end{aligned}
\tag{1.11}
$$

Theorem 1.2 (The c_r-inequality)**.** *Let $r > 0$. Suppose that $\mathbb{E}|X|^r < \infty$ and $\mathbb{E}|Y|^r <$*

∞. *Then*

$$\mathbb{E}|X+Y|^r < c_r(\mathbb{E}|X|^r + \mathbb{E}|Y|^r), \tag{1.12}$$

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

Corollary 1.3. *Suppose that* $\mathrm{Var}(X) < \infty$ *and* $\mathrm{Var}(Y) < \infty$. *Then*

$$\mathrm{Var}(X+Y) < 2(\mathrm{Var}(X) + \mathrm{Var}(Y)). \tag{1.13}$$

Proof. This follows immediately by applying Theorem 1.2 to the centered random variables $X' = X - \mathbb{E}[X]$ and $Y' = Y - \mathbb{E}[Y]$. $\square$

From (1.11) and corollary 1.3,

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

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

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

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

with

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

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

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

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

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

For the second term in (1.14) we again apply corollary 1.3:

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

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: ... Maybe finish this up later? □

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."

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:

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

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

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:

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*

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

1.5 Main Theorem

Generalization of Theorem 5.5:

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

$$\mathbb{E}[T'-T|T] = -\lambda(T-R)$$

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

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

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

$$f'(w) - wf(w) = h_{z,\alpha}(w) - \Phi(z) \tag{1.24}$$

for the smoothed indicator

$$h_{z,\alpha}(w) = \begin{cases} 1 & w \leq z \\ 1 + \frac{z-w}{\alpha} & z < w \leq z + \alpha \\ 0 & w > z + \alpha. \end{cases} \tag{1.25}$$

Therefore,

$$\begin{aligned} |P(W \leq z) - \Phi(z)| &= |\mathbb{E}[(f'(W) - Wf(W))]| \\ &= \left|\mathbb{E}\left[f'(W) - \frac{(W'-W)(f(W') - f(W))}{2\lambda} + f(W)R\right]\right| \\ &= \left|\mathbb{E}\left[f'(W)\left(1 - \frac{(W'-W)^2}{2\lambda}\right)\right.\right. \\ &\quad \left.\left. + \frac{f'(W)(W'-W)^2 - (f(W') - f(W))(W'-W)}{2\lambda} + f(W)R\right]\right| \\ &:= |\mathbb{E}[J_1 + J_2 + J_3]| \\ &\leq |\mathbb{E}[J_1]| + |\mathbb{E}[J_2]| + |\mathbb{E}[J_3]|. \end{aligned} \tag{1.26}$$

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

$$|\mathbb{E}[J_3]| \leq \mathbb{E}[|J_3|] = \mathbb{E}[|f(W)R|] \leq \mathbb{E}[|R|] \tag{1.27}$$

and

$$\begin{aligned}
|\mathbb{E}[J_1]| &= \left|\mathbb{E}\left[f'(W)\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right]\right| \\
&\leq \mathbb{E}\left[\left|f'(W)\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right|\right] \\
&\leq \mathbb{E}\left[\left|\left(1-\frac{(W'-W)^2}{2\lambda}\right)\right|\right] \\
&= \frac{1}{2\lambda}\mathbb{E}[|2\lambda - \mathbb{E}[(W'-W)^2|W]|] \\
&= \frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}[W^2]-\mathbb{E}[WR]) - \mathbb{E}[(W'-W)^2|W] + 2\lambda(1-\mathbb{E}[W^2]+\mathbb{E}[WR])|] \\
&\leq \frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}[W^2]-\mathbb{E}[WR]) - \mathbb{E}[(W'-W)^2|W]|] + \mathbb{E}[|(1-\mathbb{E}[W^2]+\mathbb{E}[WR])|]
\end{aligned} \tag{1.28}$$

Note that

$$\mathbb{E}[\mathbb{E}[(W'-W)^2|W]] = \mathbb{E}[\Delta^2] = 2\lambda(\mathbb{E}[W^2]-\mathbb{E}[WR]), \tag{1.29}$$

so

$$\frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}[W^2]-\mathbb{E}[WR])-\mathbb{E}[(W'-W)^2|W]|] \leq \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])}. \tag{1.30}$$

Combining with (1.28),

$$\begin{aligned}
|\mathbb{E}[J_1]| &\leq \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}[|1-\mathbb{E}[W^2]+\mathbb{E}[WR]|] \\
&\leq \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}[|1-\mathbb{E}[W^2]|] + \mathbb{E}[|WR|] \\
&\leq \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + |1-\mathbb{E}[W^2]| + \sqrt{\mathbb{E}[W^2]\mathbb{E}[R^2]}.
\end{aligned} \tag{1.31}$$

Lastly, we bound the second term,

$$\begin{aligned} J_2 &= \frac{1}{2\lambda}(W'-W)\int_W^{W'}(f'(W)-f'(t))dt \\ &= \frac{1}{2\lambda}(W'-W)\int_W^{W'}\int_t^W f''(u)dudt \\ &= \frac{1}{2\lambda}(W'-W)\int_W^{W'}(W'-u)f''(u)du. \end{aligned} \tag{1.32}$$

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,

$$
\begin{aligned}
|\mathbb{E}[J_2]| &= \left|\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)\int_W^{W'}(W'-u)f''(u)du\right]\right| \\
&= \left|\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)\int_W^{W'}\left(\frac{W+W'}{2}-u\right)f''(u)du\right]\right| \\
&\leq \left|\mathbb{E}\left[||f''||\frac{1}{2\lambda}|W'-W|\int_{\min(W,W')}^{\max(W,W')}\left|\frac{W+W'}{2}-u\right|du\right]\right| \qquad (1.33)\\
&= \left|\mathbb{E}\left[||f''||\frac{1}{2\lambda}\frac{|W'-W|^3}{4}\right]\right| \\
&\leq \frac{\mathbb{E}[|W'-W|^3]}{4\alpha\lambda},
\end{aligned}
$$

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

$$
\begin{aligned}
P(W \leq z) &\leq \mathbb{E}[h_{z,\alpha}(W)] \\
&\leq Nh_{z,\alpha} + \frac{\mathbb{E}[|W'-W|^3]}{4\alpha\lambda} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} \\
&\quad + |1-\mathbb{E}[W^2]| + \sqrt{\mathbb{E}[W^2]\mathbb{E}[R^2]} + \mathbb{E}[|R|] \qquad (1.34)\\
&\leq \Phi(z) + \frac{\alpha}{\sqrt{2\pi}} + \frac{\mathbb{E}[|W'-W|^3]}{4\alpha\lambda} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} \\
&\quad + |1-\mathbb{E}[W^2]| + \sqrt{\mathbb{E}[W^2]\mathbb{E}[R^2]} + \mathbb{E}[|R|]
\end{aligned}
$$

The minimizer of the expression is

$$
\alpha = \frac{(2\pi)^{1/4}}{2}\sqrt{\frac{\mathbb{E}[|W'-W|^3]}{\lambda}}. \qquad (1.35)
$$

Plugging this in, we get the upper bound

$$
\begin{aligned}
P(W \leq z) - \Phi(z) \leq (2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}[|W'-W|^3]}{\lambda}} + \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} \qquad (1.36)\\
+ |1-\mathbb{E}[W^2]| + \sqrt{\mathbb{E}[W^2]\mathbb{E}[R^2]} + \mathbb{E}[|R|]
\end{aligned}
$$

Proving the corresponding lower bound in a similar manner completes the proof of the theorem. $\square$

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, using the theorem of chapter 1.

2.1 Set-up

We observe two samples with equal sample size: $\{u_i\}_{i=1}^N$ and $\{u_i\}_{i=N+1}^{2N}$.

Student's two-sample t-statistic is given by

$$\begin{aligned}
T(\{u_i\}_{i=1}^N, \{u_i\}_{i=N+1}^{2N}) &= \frac{\bar{u}_1 - \bar{u}_2}{\sqrt{\frac{\frac{1}{N-1}\sum_{i=1}^N (u_i - \bar{u}_1)^2}{N} + \frac{\frac{1}{N-1}\sum_{i=N+1}^{2N}(u_i - \bar{u}_2)^2}{N}}} \\
&= \frac{1}{\sqrt{\frac{N}{N-1}}} \frac{\sum_{i=1}^N u_i - \sum_{i=N+1}^{2N} u_i}{\sqrt{\sum_{i=1}^N (u_i - \bar{u}_1)^2 + \sum_{i=N+1}^{2N}(u_i - \bar{u}_2)^2}} \\
&= \sqrt{\frac{N-1}{N}} \frac{q}{d},
\end{aligned}$$

where

$$
\begin{aligned}
q &= \left(\sum_{i=1, i\neq I}^{N} u_i + u_I - \sum_{i=N+1, i\neq J}^{2N} u_i - u_J\right) \\
d &= \sqrt{\sum_{i=1}^{N}(u_i - \bar{u}_1)^2 + \sum_{i=N+1}^{2N}(u_i - \bar{u}_2)^2}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
T'(\{u_i\}_{i=1}^{N}, \{u_i\}_{i=N+1}^{2N}) &= \sqrt{\frac{N-1}{N}}\frac{q'}{d'} \\
q' &= \left(\sum_{i=1, i\neq I}^{N} u_i + u_J - \sum_{i=N+1, i\neq J}^{2N} u_i - u_I\right) \\
&= q - 2u_I + 2u_J \\
d' &= \sqrt{\sum_{i=1}^{N}(u_i - \bar{u}_1')^2 + \sum_{i=N+1}^{2N}(u_i - \bar{u}_2')^2}.
\end{aligned}
$$

In order to perform hypothesis testing, we would like to know the randomization distribution of T. We shall create an exchangeable pair (T, T') 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 $T' = T$, where T' is the t-statistic under this random transposition. That is, the t-statistic is invariant to within-group transpositions. Thus, 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 $N+1 \leq J \leq 2N$.

2.2 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 proper transformation is

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

so we just consider the u_i's as having been transformed. This can be seen as a very mild assumption of disallowing the case where all our data are constant.

We also assume that

$$B = \max_{\pi} \bar{u}_2^2 < 1. \tag{2.2}$$

This rejects situations like $(1, 1, \ldots, -1, -1)$.

2.3 Preliminaries

Here we collect useful bounds and other results.

In order to bound various moments of $\bar{u}_2$ under the permutation distribution, we use a result of Serfling's [1]:

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

$$\begin{aligned}
\mathbb{E}[\bar{u}_2^p] &\leq \frac{\Gamma(p/2+1)}{2^{p/2+1}} \left[\frac{N+1}{2N}(b-a)^2\right]^{p/2} (2N)^{-p/2} \\
&\leq \frac{\Gamma(p/2+1)}{2^{p/2+1}} \left[\frac{N+1}{4N}(b-a)^2\right]^{p/2} (N)^{-p/2} \\
&\leq \frac{\Gamma(p/2+1)}{2^{p/2+1}} \left[\frac{1}{2}(b-a)^2\right]^{p/2} N^{-p/2} \\
&:= f_{c_1}(p) N^{-p/2}.
\end{aligned} \tag{2.3}$$

By assumption,

$$\begin{aligned}
d^{-p} &= \frac{1}{(2N(1-\bar{u}_2^2))^{p/2}} \\
&\le \frac{1}{(2N(1-B^2))^{p/2}} \\
&= \frac{1}{(2(1-B^2))^{p/2}} N^{-p/2} \\
&:= f_{c_2}(p) N^{-p/2}. \qquad (2.4)
\end{aligned}$$

The transposition (I, J) also affects the denominator of T', and we need to quantify the difference between the denominators of T and T'.

$$\begin{aligned}
d^2 &= \sum_{i=1}^{N}(u_i - \bar{u}_1)^2 + \sum_{i=N+1}^{2N}(u_i - \bar{u}_2)^2 = \sum_{i=1}^{2N} u_i^2 - N\bar{u}_1^2 - N\bar{u}_2^2 \\
d'^2 &= \sum_{i=1}^{2N} u_i^2 - N\bar{u}_1'^2 - N\bar{u}_2'^2,
\end{aligned}$$

where

$$\bar{u}_1' = \bar{u}_1 - \frac{1}{N}u_I + \frac{1}{N}u_J \text{ and } \bar{u}_2' = \bar{u}_2 - \frac{1}{N}u_J + \frac{1}{N}u_I.$$

So,

$$\bar{u}_1'^2 = \bar{u}_1^2 + \frac{2\bar{u}_1}{N}(u_J - u_I) + \frac{1}{N^2}(u_J - u_I)^2$$

and

$$\bar{u}_2'^2 = \bar{u}_2^2 + \frac{2\bar{u}_2}{N}(u_I - u_J) + \frac{1}{N^2}(u_I - u_J)^2.$$

Since $\sum u_i = 0$, $\bar{u}_1 = -\bar{u}_2$, so

$$\begin{aligned}
h &= d^2 - d'^2 \\
&= -N\bar{u}_1^2 - N\bar{u}_2^2 + N\bar{u}_1'^2 + N\bar{u}_2'^2 \\
&= 2\bar{u}_1(u_J - u_I) + 2\bar{u}_2(u_I - u_J) + \frac{2}{N}(u_I - u_J)^2 \\
&= 4\bar{u}_2(u_I - u_J) + \frac{2}{N}(u_I - u_J)^2
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}[h^p] &= \mathbb{E}\left[\left|4\bar{u}_2(u_I-u_J)+\frac{2}{N}(u_I-u_J)^2\right|^p\right] \\
&\le 2^{p-1}\left(\mathbb{E}[|4\bar{u}_2(u_I-u_J)|^p]+\mathbb{E}\left[\left|\frac{2}{N}(u_I-u_J)^2\right|^p\right]\right) \\
&\le 2^{p-1}\left[(4(b-a))^p\mathbb{E}\,|\bar{u}_2|^p+\left(\frac{2}{N}(b-a)^2\right)^p\right] \\
&\le 2^{p-1}(4(b-a))^p f_{c_1}(p)N^{-p/2}+2^{p-1}(2(b-a)^2)^pN^{-p/2}N^{-p/2} \\
&\le (2^{p-1}(4(b-a))^p f_{c_1}(p)N^{-p/2}+2^{p-1}(2(b-a)^2)^p)N^{-p/2} \\
&:= f_{c_3}(p)N^{-p/2}. \qquad (2.5)
\end{aligned}$$

Now we consider the difference $d-d'$. For a given d (which grows linearly), consider d' as a function of the difference:

$$d'=\sqrt{d^2-h}=f(h)=f(0)+f'(0)h+\ldots=d-\frac{h}{2d}+\ldots$$

The derivative is

$$f'(h)=\frac{d}{\sqrt{d^2-h}}$$

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

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

Here, we are approximating d' by a constant and bounding the error by using the first derivative, but it's okay because the square root function flattens out and the difference inside the square root is probabilistically small.

Now

$$|d-d'|\le\frac{|h|}{2\sqrt{d^2-\xi_L}}\le\frac{|h|}{2\sqrt{d^2-\max(0,h)}}$$

Recall that $h = d^2 - d'^2$, so

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

Therefore,

$$|d - d'| \leq \frac{|h|}{2\min(d, d')} \leq \max\left(\frac{|h|}{2d}, \frac{|h|}{2d'}\right) \leq \frac{|h|}{2d} + \frac{|h|}{2d'}.$$

The important thing to do is to isolate $|h|$, which is small in expectation, but not absolutely.

$$\begin{aligned}
\mathbb{E}|d - d'|^p &\leq 2^{p-1}\left(\mathbb{E}\left|\frac{h}{2d}\right|^p + \mathbb{E}\left|\frac{h}{2d'}\right|^p\right) \\
&\leq 2^{-1}\left(\mathbb{E}\left|\frac{h}{d}\right|^p + \mathbb{E}\left|\frac{h}{d'}\right|^p\right) \\
&\leq 2^{-1}(\sqrt{\mathbb{E}[h^{2p}]\mathbb{E}[d^{-2p}]} + \sqrt{\mathbb{E}[h^{2p}]\mathbb{E}[d'^{-2p}]}) \\
&\leq \sqrt{\mathbb{E}[h^{2p}]\mathbb{E}[d^{-2p}]} \\
&\leq \sqrt{f_{c_3}(2p)N^{-2p/2} f_{c_2}(2p)N^{-2p/2}} \\
&\leq \sqrt{f_{c_3}(2p)f_{c_2}(2p)}N^{-p} \\
&:= f_{c_4}(p)N^{-p}.
\end{aligned} \tag{2.6}$$

With $q = N\bar{u}_1 - N\bar{u}_2 = -2N\bar{u}_2$, and noting that q and q' are exchangeable,

$$\begin{aligned}
\mathbb{E}[q'^p] &= \mathbb{E}[q^p] \\
&= \mathbb{E}[(-2N\bar{u}_2)^p] \\
&= (-2N)^p\mathbb{E}[\bar{u}_2^p] \\
&\leq 2^p N^p f_{c_1}(p)N^{-p/2} \text{ from (2.3)} \\
&= 2^p f_{c_1}(p)N^{p/2} \\
&:= f_{c_5}(p)N^{p/2}.
\end{aligned} \tag{2.7}$$

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{q'}{dd'}\right)^p\right] &\leq \sqrt{\mathbb{E}|q'|^{2p}\mathbb{E}|dd'|^{-2p}} \\
&\leq \sqrt{\mathbb{E}|q|^{2p}\sqrt{\mathbb{E}|d|^{-4p}\mathbb{E}|d'|^{-4p}}} \\
&\leq \sqrt{\mathbb{E}|q|^{2p}\sqrt{\mathbb{E}|d|^{-4p}\mathbb{E}|d|^{-4p}}} \\
&\leq \sqrt{\mathbb{E}|q|^{2p}\mathbb{E}|d|^{-4p}} \\
&\leq \sqrt{f_{c_5}(2p)N^{2p/2}f_{c_2}(4p)N^{-4p/2}} \text{ from (2.7) and (2.4)} \\
&\leq \sqrt{f_{c_5}(2p)f_{c_2}(4p)}N^{-p/2} \\
&:= f_{c_6}(p)N^{-p/2}. \qquad (2.8)
\end{aligned}
$$

References

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