TOPICS IN TWO-SAMPLE TESTING

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

Nelson C. Ray
2013

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

(Susan P. Holmes) Principal Adviser

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

(Persi W. Diaconis)

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

(Bradley Efron)

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

(Jerome H. Friedman)

Approved for the University Committee on Graduate Studies.

Contents

Chapter 1

Stein's method

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

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

1.1 Introduction

Stein's method provides a means of bounding the distance between two probability distributions in a given probability metric. When applied with the normal distribution as the target, this results in central-limit-type theorems. Several flavors of Stein's method (e.g. the method of exchangeable pairs) proceed via auxiliary randomization. In Appendix B, we reproduce Stein's [3] proof of the HCCLT.

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

1.2 Stein's Theorem

Theorem 1.1 bounds the Kolmogorov distance between the distribution of the random variable W and the standard normal distribution in terms of functions of the difference of the exchangeable pair (W, W'). It is applied to the situation where W is the sum of the random diagonal of a matrix to prove the HCCLT. Chen et al. generalize Theorem 1.1 to allow for situations in which the regression condition does not hold exactly, and we later in addition relax the assumption that the variance is 1.

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

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

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

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

1.3 Hoeffding's Combinatorial CLT

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

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

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

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

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

Define

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

to be the sum of a random diagonal. Then

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

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

Bolthausen [1] was able to prove the following result:

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

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

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

Now, we return to generalizing Theorem 1.1.

1.4 Generalized Stein's Theorems

Here, we treat the situation where the regression condition fails to hold exactly. Chen et al. [2] serves as an excellent reference for results of this type.

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

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

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

Here we generalize Lemma 5.1 from [2] to setting of non-unit variance:

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

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

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

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

and

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

This theorem will allow us to prove a $\mathcal{O}(N^{-1/4})$ rate under mild conditions.

Generalization of Theorem 5.5 from [2]:

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

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

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

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

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

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

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

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

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

Appendix A

Auxiliary Results

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

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

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

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

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

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

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

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

Here is a slight generalization of Lemma 2.7 from [2]:

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

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

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

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

Proof. From (1.1) we have

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

We also have

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

using exchangeability. Equating the two expressions yields

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

As an intermediate computation,

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

Then

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

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

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

Rearranging the expression yields

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

□

This is just a small part of Lemma 2.4 from [2]:

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

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

Lemma 2.2 from [2]:

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

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

is given by

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

Part of Lemma 2.3 from [2]:

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

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

Appendix B

Stein's Method Proofs

B.0.1 Proof of Theorem 1.2

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

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

Define the random permutation Π' by

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

We construct our exchangeable pair by defining

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

We now verify the contraction property:

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

This satisfies our contraction property with

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

To bound the variance component, compute

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

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

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

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

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

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

with

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

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

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

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

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

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

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

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

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

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

B.0.2 Proof of Lemma 1.5

Proof. Let

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

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

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

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:

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

□

B.0.3 Proof of Theorem 1.6

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

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

for the smoothed indicator

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

Therefore,

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

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

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

and

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

Note that

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

so

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

Combining with (B.13),

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

Lastly, we bound the second term,

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

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

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

For the latter,

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

Since W and W' are exchangeable,

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

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

Collecting the bounds, we obtain

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

The minimizer of the expression is

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

Plugging this in, we get the upper bound

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

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

B.0.4 Proof of Theorem 1.7

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

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

Using (A.9), $f'(W) = Wf(W) + \mathbf{1}_{\{w\leq z\}} - \Phi(z)$ 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 (A.7) 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

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

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

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

Therefore,

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

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

References

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

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

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