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 $\{1,\ldots,n\}$ 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:

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

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:

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

This satisfies our contraction property with

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

To bound the variance component, compute

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

From (1.11) and corollary 1.3,

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

for $n \geq 2$ 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

$$\sum_{i,j=1,i\neq j}^{n} \text{Cov}(a_{i\Pi(i)}^2, a_{j\Pi(j)}^2) = \sum_{i,j=1,i\neq j}^{n} \left(\frac{1}{n(n-1)} \sum_{k,l=1,k\neq l}^{n} a_{ik}^2 a_{jl}^2 - \left(\frac{1}{n} \sum_{k} a_{ik}^2 \right) \left(\frac{1}{n} \sum_{l} a_{jl}^2 \right) \right)$$

$$= \sum_{i,j=1,i\neq j}^{n} \left(\frac{1}{n(n-1)} \sum_{k,l=1}^{n} a_{ik}^2 a_{jl}^2 - \frac{1}{n^2} \sum_{k} \sum_{l} a_{ik}^2 a_{jl}^2 - \frac{1}{n(n-1)} \sum_{k} a_{ik}^2 a_{jk}^2 \right)$$

$$= \frac{1}{n^2(n-1)} \sum_{i,j=1,i\neq j}^{n} \sum_{k,l=1}^{n} a_{ik}^2 a_{jl}^2 - \frac{1}{n(n-1)} \sum_{i,j=1,i\neq j}^{n} \sum_{k} a_{ik}^2 a_{jk}^2$$

$$\leq \frac{(n-1)^2}{n^2(n-1)}$$

$$\leq \frac{1}{n}$$

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,

$$(n-1)^2 = (\mathbf{a}^T \mathbf{1})^2$$

$$\leq ||\mathbf{a}||_2^2 ||\mathbf{1}||_2^2$$

$$= n^2 \sum_{i,j=1}^{n} a_{i,j}^4.$$

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

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

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

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

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? $\square$

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.

1.3.1 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 antisymmetric 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*

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

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

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

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,

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

Then

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

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

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

$\square$

Generalization of Theorem 5.5:

Theorem 1.7. *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

$$
\sup_{t \in \mathbb{R}} |P(T \leq t) - \Phi(t)| \leq B + (2\pi)^{-1/4}\sqrt{\frac{\mathbb{E}|T' - T|^3}{\lambda}} + \mathbb{E}|R|,
$$

where $B \leq \frac{\Theta}{2\lambda}$ and $\Theta = \sqrt{\mathrm{Var}(\mathbb{E}[(T' - T)^2|T])}$.

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

$$
f'(w) - wf(w) = h_{z,\alpha}(w) - \Phi(z) \tag{1.23}
$$

for the smoothed indicator

$$
h_{z,\alpha}(w) = \begin{cases} 1 & w \leq z \\ 1 + \frac{z - w}{\alpha} & z < w \leq z + \alpha \\ 0 & w > z + \alpha. \end{cases} \tag{1.24}
$$

Therefore,

$$|P(W \le z) - \Phi(z)| = |\mathbb{E}[(f'(W) - Wf(W))]|$$

$$= \left| \mathbb{E}\left[f'(W) - \frac{(W'-W)(f(W')-f(W))}{2\lambda} + f(W)R \right] \right|$$

$$= \left| \mathbb{E}\left[f'(W)\left(1 - \frac{(W'-W)^2}{2\lambda} \right) \right.\right.$$

$$\left.\left. + \frac{f'(W)(W'-W)^2 - (f(W')-f(W))(W'-W)}{2\lambda} + f(W)R \right] \right|$$

$$:= |\mathbb{E}[J_1 + J_2 + J_3]|$$

$$\le |\mathbb{E}[J_1]| + |\mathbb{E}[J_2]| + |\mathbb{E}[J_3]|.$$

$$(1.25)$$

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

$$|\mathbb{E}[J_3]| \le \mathbb{E}[|J_3|] = \mathbb{E}[|f(W)R|] \le \mathbb{E}[|R|] \qquad (1.26)$$

and

$$|\mathbb{E}[J_1]| = \left| \mathbb{E}\left[f'(W)\left(1 - \frac{(W'-W)^2}{2\lambda} \right) \right] \right|$$

$$\le \mathbb{E}\left[\left| f'(W)\left(1 - \frac{(W'-W)^2}{2\lambda} \right) \right| \right]$$

$$\le \mathbb{E}\left[\left| \left(1 - \frac{(W'-W)^2}{2\lambda} \right) \right| \right]$$

$$= \frac{1}{2\lambda}\mathbb{E}[|2\lambda - \mathbb{E}[(W'-W)^2|W]|]$$

$$= \frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}[W^2] - \mathbb{E}[WR]) - \mathbb{E}[(W'-W)^2|W] + 2\lambda(1 - \mathbb{E}[W^2] + \mathbb{E}[WR])|]$$

$$\le \frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}[W^2] - \mathbb{E}[WR]) - \mathbb{E}[(W'-W)^2|W]|] + \mathbb{E}[|(1 - \mathbb{E}[W^2] + \mathbb{E}[WR])|]$$

$$(1.27)$$

Note that

$$\mathbb{E}[\mathbb{E}[(W'-W)^2|W]] = \mathbb{E}[\Delta^2] = 2\lambda(\mathbb{E}[W^2] - \mathbb{E}[WR]), \qquad (1.28)$$

so

$$\frac{1}{2\lambda}\mathbb{E}[|2\lambda(\mathbb{E}[W^2]-\mathbb{E}[WR])-\mathbb{E}[(W'-W)^2|W]|] \leq \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])}. \quad (1.29)$$

Combining with (1.27),

$$\begin{aligned}
|\mathbb{E}[J_1]| &\leq \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}[|1-\mathbb{E}[W^2]+\mathbb{E}[WR]|] \\
&\leq \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}[|1-\mathbb{E}[W^2]|] + \mathbb{E}[|WR|] \quad (1.30) \\
&\leq \frac{1}{2\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W'-W)^2|W])} + \mathbb{E}[|1-\mathbb{E}[W^2]|] + \sqrt{\mathbb{E}[W^2]\mathbb{E}[R^2]}.
\end{aligned}$$

Lastly, we bound the second term,

$$\begin{aligned}
J_2 &= \frac{1}{2\lambda}(W'-W)\int_W^{W'}(f'(W)-f'(t))dt \\
&= \frac{1}{2\lambda}(W'-W)\int_W^{W'}\int_t^W f''(u)dudt \quad (1.31) \\
&= \frac{1}{2\lambda}(W'-W)\int_W^{W'}(W'-u)f''(u)du.
\end{aligned}$$

$\square$