Stein's method using approximate zero bias couplings with applications to combinatorial central limit theorems under the Ewens distribution

Nathakhun Wiroonsri

Department of Mathematics, University of Southern California
3620 S. Vermont Ave., KAP 104,
Los Angeles, CA 90089, USA.
E-mail address: `wiroonsr@usc.edu`

Abstract. We generalize the well-known zero bias distribution and the λ-Stein pair to an approximate zero bias distribution and an approximate λ, R-Stein pair, respectively. Berry Esseen type bounds to the normal, based on approximate zero bias couplings and approximate λ, R-Stein pairs, are obtained using Stein's method. The bounds are then applied to combinatorial central limit theorems where the random permutation has the Ewens $\mathcal{E}_\theta$ distribution with $\theta > 0$ which can be specialized to the uniform distribution by letting $\theta = 1$. The family of the Ewens distributions appears in the context of population genetics in biology.

1. Introduction

We develop L^1 and L^∞ bounds for normal approximation using Stein's method, based on approximate zero bias couplings. The results are applied to the combinatorial central limit theorems, that is, we derive such bounds for the distribution, introduced in Hoeffding (1951), of

$$Y = \sum_{i=1}^{n} a_{i,\pi(i)} \tag{1.1}$$

where $A \in \mathbb{R}^{n \times n}$ is a given real matrix with components $\{a_{i,j}\}_{i,j=1}^{n}$ and $\pi \in \mathcal{S}_n$ has the Ewens distribution. We recall that the L^1 and L^∞ distances between the distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ of real valued random variables X and Y are given,

respectively, by

$$
\begin{aligned}
d_1\big(\mathcal{L}(X), \mathcal{L}(Y)\big) &= \int_{-\infty}^{\infty} |P(X \le t) - P(Y \le t)| dt \\
&= \sup_{h \in \mathcal{H}_1} |\mathbb{E}h(X) - \mathbb{E}h(Y)| \tag{1.2}
\end{aligned}
$$

where $\mathcal{H}_1 = \{h : |h(y) - h(x)| \le |y - x|\}$, and

$$
\begin{aligned}
d_\infty\big(\mathcal{L}(X), \mathcal{L}(Y)\big) &= \sup_{\infty < t < \infty} |P(X < t) - P(Y < t)| \\
&= \sup_{h \in \mathcal{H}_\infty} |\mathbb{E}h(X) - \mathbb{E}h(Y)| \tag{1.3}
\end{aligned}
$$

where $\mathcal{H}_\infty = \{\mathbf{1}(\cdot \le t) : t \in \mathbb{R}\}$. In the following, we will drop the subscripts 1 and ∞ when the statement is true for both $\mathcal{H}_1$ and $\mathcal{H}_\infty$.

Stein's method for normal approximation, introduced in Stein (1972) (see also the text Chen et al., 2011 and the introductory notes Ross, 2011), was motivated from the fact that W has the standard normal distribution, denoted $\mathcal{N}(0,1)$, if and only if

$$
\mathbb{E}W f(W) = \mathbb{E}f'(W)
$$

for all absolutely continuous functions f with $\mathbb{E}|f'(W)| < \infty$. This equation and the form of the distances in (1.2) and (1.3) lead to the differential equation

$$
h(w) - Nh = f_h'(w) - w f_h(w) \tag{1.4}
$$

where $Nh = \mathbb{E}h(Z)$ with $Z \sim \mathcal{N}(0,1)$ and $h \in \mathcal{H}$. Taking the supremum over all $h \in \mathcal{H}_1$ (resp. $h \in \mathcal{H}_\infty$) to the expectation on the left hand side of (1.4) with w replaced by a variable W yields the distance between W and Z in (1.2) (resp. (1.3)). Thus, instead of working on the distances directly, one can handle the expectation on the right hand side using the bounded solution f_h of (1.4) for the given h. Using this device, Stein's method has uncovered an alternative way to show convergence in distribution with additional information on the finite sample distance between distributions and can also deal with various kinds of dependence through the help of coupling constructions.

One of the well-known couplings in the literature is the *zero bias coupling* which was first introduced in Goldstein and Reinert (1997). Recall that for X with mean zero and variance $\sigma^2 \in (0, \infty)$, we say that X^z has the X-zero biased distribution if

$$
\mathbb{E}X f(X) = \sigma^2 \mathbb{E}f'(X^z) \tag{1.5}
$$

for all absolutely continuous functions f for which the expectations exist. Applying (1.5) in (1.4) with X replaced by X/σ, we have

$$
\begin{aligned}
\mathbb{E}h(X/\sigma) - Nh &= \mathbb{E}\left(f'\left(\frac{X}{\sigma}\right) - \frac{X}{\sigma}f\left(\frac{X}{\sigma}\right)\right) \\
&= \mathbb{E}\left(f'\left(\frac{X}{\sigma}\right) - f'\left(\frac{X^z}{\sigma}\right)\right). \tag{1.6}
\end{aligned}
$$

Generalizing the proofs of Theorems 4.1 and 5.1 of Chen et al. (2011) that only showed the results for $\sigma^2 = 1$, with the help of (1.6), we have

$$
d_1(\mathcal{L}(X/\sigma), \mathcal{L}(Z)) \le \frac{2}{\sigma}\mathbb{E}|X^z - X|, \tag{1.7}
$$

and, with $|X^z - X| \le \delta$,

$$d_\infty(\mathcal{L}(X/\sigma), \mathcal{L}(Z)) \le \frac{(1 + 1/\sqrt{2\pi} + \sqrt{2\pi}/4)\delta}{\sigma}. \tag{1.8}$$

It is easy to see from (1.7) and (1.8) that once the zero bias coupling has been constructed in such a way that the two variables are close, one can simply obtain good bounds for L^1 and L^∞ distances. Nevertheless, the difficult part is that there is no general way to construct the zero bias coupling. One of the most efficient method, introduced in Goldstein and Reinert (1997), is to take advantage of the existence of a λ-Stein pair. We recall that an exchangeable pair X', X'' forms a λ-Stein pair if

$$\mathbb{E}(X''|X') = (1 - \lambda)X'$$

for some $0 < \lambda < 1$. The following lemma illustrates the way to construct zero bias couplings through λ-Stein pairs.

Lemma 1.1 (Goldstein and Reinert, 1997). *Let X', X'' be a λ-Stein pair with* $\mathrm{Var}(X') = \sigma^2 \in (0, \infty)$ *and distribution $F(x', x'')$. Then when $X^\dagger, X^\ddagger$ have distribution*

$$dF^\dagger(x', x'') = \frac{(x' - x'')^2}{\mathbb{E}(X' - X'')^2} dF(x', x''),$$

and U be uniform $\mathcal{U}[0, 1]$ and is independent of $X^\dagger, X^\ddagger$, the variable

$$X^z = UX^\dagger + (1 - U)X^\ddagger \quad \text{has the } X'\text{-zero biased distribution.}$$

Lemma 1.1 has been used in several works. Berry Esseen type bounds to the normal, based on zero bias couplings, were first obtained in Goldstein (2005). The bounds were then applied to the combinatorial central limit theorems where the random permutation has either the uniform distribution or one which is constant over permutations with the same cycle type and having no fixed points. Concentration inequalities on the same setting were shown in Goldstein and I̧slak (2014). In Fulman and Goldstein (2011), the zero biasing and this lemma were also used to obtain L^1 bounds to the normal for a variable constructed from an interesting property of the Jack_α measure on the set of partitions of size n. Apart from the use of this lemma, Stein's method with different techniques has been applied to the combinatorial central limit theorems under the uniform distribution in several works. One of the most recent papers is Chen and Fang (2015) where the exchangeable pairs technique was used to obtain L^∞ bounds between Y as in (1.1) with a fixed matrix A replaced by independent random variables $\{X_{i,j} : i, j \in [n]\}$ and the normal distribution. Here for a positive integer m we denote $[m] = \{1, 2, \ldots, m\}$. The bounds there are given in term of the third moments of $X_{i,j}$. To learn more about the history of Stein's method and the combinatorial central limit theorems, see the references therein. Without using Stein's method, the results were generalized to the case without third moments in Frolov (2014) and to various moment conditions in Frolov (2017). The original idea of replacing $a_{i,j}$ by $X_{i,j}$ dates back to Ho and Chen (1978) where the results are optimal only in the case that there exists $C > 0$ such that $|X_{i,j}| \le C$ for all $i, j \in [n]$.

Although the combinatorial central limit theorems have attracted attention for quite some time and have been extended to different settings, the case where the

random permutation has the Ewens distribution has yet been studied. It is inter-
esting to investigate the robustness and the sensitivity of normality when the usual
assumptions in the uniform distribution or the distribution that is constant over
permutations with the same cycle type and having no fixed points are not satisfied.
This is useful in the real-life situation as the uniform properties are sometimes be-
lieved to hold but actually do not. One difficulty that may arise in order to use
Lemma 1.1 is that a λ-Stein pair does not always exist. However, one might be able
to construct a pair which has nearly the same condition as the λ-Stein pair and
this is where we start. We call a pair of random variables (Y', Y''), an *approximate*
λ, R-*Stein pair* if it is exchangeable and satisfies

$$\mathbb{E}Y' = 0, \quad \mathrm{Var}(Y') = \sigma^2$$

with $\sigma^2 \in (0, \infty)$, and

$$\mathbb{E}(Y''|Y') = (1 - \lambda)Y' + R, \tag{1.9}$$

for some $0 < \lambda < 1$ and $R = R(Y')$.

In this work, we generalize Lemma 1.1 to a new version, Lemma 2.1, replacing a
λ-Stein pair by an approximate λ, R-Stein pair. The lemma leads to a variable that
is similar to the zero bias variable but has two extra terms depending on R and we
call it an *approximate zero bias variable*. Then we also generalize the L^1 and L^∞
bounds in (1.7) and (1.8) using approximate zero bias couplings in Theorems 2.4
and 2.5, respectively.

The remainder of this work is organized as follows. In Section 2 we state and
prove the general results for approximate λ, R-Stein pairs and approximate zero bias
variables. These results are then applied, in Section 4, to the combinatorial central
limit theorems where the random permutation has the Ewens distribution. The
description of the Ewens distribution and some necessary properties are presented
in Section 3. In Section 5, the Appendix, we prove some of the results from Section 4
that are straightforward but requires some attention to detail.

2. Main results: approximate λ, R-Stein pairs and approximate zero bias couplings

Let (Y', Y'') be an approximate λ, R-Stein pair. Taking expectation in (1.9),
using exchangeability and that Y' has mean zero yields

$$\mathbb{E}R = (1 - \lambda)\mathbb{E}Y' + \mathbb{E}R = \mathbb{E}Y'' = 0.$$

In addition, for any function f such that the following expectations exist,

$$\begin{aligned}
\mathbb{E}Y''f(Y') &= \mathbb{E}(\mathbb{E}(Y''f(Y')|Y')) = \mathbb{E}(f(Y')\mathbb{E}(Y''|Y')) \\
&= \mathbb{E}(f(Y')((1 - \lambda)Y' + R)) \\
&= (1 - \lambda)\mathbb{E}Y'f(Y') + \mathbb{E}Rf(Y'). \tag{2.1}
\end{aligned}$$

In particular, specializing (2.1) to the case $f(y) = y$ yields

$$\mathbb{E}Y''Y' = (1 - \lambda)\mathbb{E}Y'^2 + \mathbb{E}Y'R,$$

and thus

$$\begin{aligned}
\mathbb{E}(Y'' - Y')^2 &= 2\left(\mathbb{E}Y'^2 - \mathbb{E}Y''Y'\right) \\
&= 2\left(\lambda\mathbb{E}Y'^2 - \mathbb{E}Y'R\right) = 2\left(\lambda\sigma^2 - \mathbb{E}Y'R\right). \tag{2.2}
\end{aligned}$$

Now we state and prove the following lemma which is the generalized version of Lemma 1.1 adapted to approximate λ, R-Stein pairs. We call a variable Y^* that satisfies (2.3) below an *approximate Y'-zero bias variable*.

Lemma 2.1. *Let (Y', Y'') be an approximate λ, R-Stein pair with distribution $F(y', y'')$. Then when $(Y^\dagger, Y^\ddagger)$ has distribution*

$$dF^\dagger(y', y'') = \frac{(y'' - y')^2}{\mathbb{E}(Y'' - Y')^2} dF(y', y''),$$

and $U \sim \mathcal{U}([0,1])$ is independent of $Y^\dagger, Y^\ddagger$, the variable $Y^* = UY^\dagger + (1 - U)Y^\ddagger$ satisfies

$$\mathbb{E}[Y'f(Y')] = \sigma^2 \mathbb{E}f'(Y^*) - \frac{\mathbb{E}Y'R}{\lambda}\mathbb{E}f'(Y^*) + \frac{\mathbb{E}Rf(Y')}{\lambda} \qquad (2.3)$$

for all absolutely continuous functions f.

Proof: For all absolutely continuous functions f for which the expectations below exist,

$$
\begin{aligned}
\sigma^2 \mathbb{E}f'(Y^*) &= \sigma^2 \mathbb{E}f'(UY^\dagger + (1-U)Y^\ddagger) \\
&= \sigma^2 \mathbb{E}\left(\frac{f(Y^\dagger) - f(Y^\ddagger)}{Y^\dagger - Y^\ddagger}\right) \\
&= \frac{\sigma^2}{2(\lambda\sigma^2 - \mathbb{E}Y'R)}\mathbb{E}\left(\left(\frac{f(Y'') - f(Y')}{Y'' - Y'}\right)(Y'' - Y')^2\right) \\
&= \frac{\sigma^2}{2(\lambda\sigma^2 - \mathbb{E}Y'R)}\mathbb{E}((f(Y'') - f(Y'))(Y'' - Y')) \\
&= \frac{\sigma^2}{\lambda\sigma^2 - \mathbb{E}Y'R}\mathbb{E}(Y'f(Y') - Y''f(Y')) \\
&= \frac{\lambda\sigma^2}{\lambda\sigma^2 - \mathbb{E}Y'R}\mathbb{E}(Y'f(Y')) - \frac{\sigma^2}{\lambda\sigma^2 - \mathbb{E}Y'R}\mathbb{E}Rf(Y'),
\end{aligned}
$$

where we have used (2.2) and (2.1) in the third and the last equalities, respectively.

Thus

$$
\begin{aligned}
\mathbb{E}Y'f(Y') &= \frac{\lambda\sigma^2 - \mathbb{E}Y'R}{\sigma^2\lambda}\sigma^2\mathbb{E}f'(Y^*) + \frac{\mathbb{E}Rf(Y')}{\lambda} \\
&= \sigma^2\mathbb{E}f'(Y^*) - \frac{\mathbb{E}Y'R}{\lambda}\mathbb{E}f'(Y^*) + \frac{\mathbb{E}Rf(Y')}{\lambda}.
\end{aligned}
$$

$\square$

Remark 2.2. One may notice that we construct an approximate zero bias coupling through an exchangeable pair. An important reason that we develop this coupling technique instead of simply using Stein's method of exchangeable pairs is that we aim to avoid the calculation of the term $\mathrm{Var}\left(\mathbb{E}[(Y'' - Y')^2|Y']\right)$ that can be difficult to compute in many cases. The reader will see an example in Section 4 that we take one more step that might not be very easy to construct an approximate zero bias coupling but all the computations after that are straightforward.

Next the following result shows how to construct an approximate zero bias distribution of Y' using an approximate Stein pair (Y', Y''), when the latter is a function of some underlying random variables $\xi_\alpha, \alpha \in \chi$ and a random index **I**. It is a minor variation of Lemma 4.4 of Chen et al. (2011), with a λ-Stein

pair and $2\lambda\sigma^2$ there respectively replaced by an approximate λ, R-Stein pair and $\mathbb{E}(Y' - Y'')^2 = 2\left(\lambda\sigma^2 - \mathbb{E}Y'R\right)$ here. The proof is omitted, being similar under these replacements.

Lemma 2.3. *Let $F(y', y'')$ be the distribution of an approximate λ, R-Stein pair (Y', Y'') and suppose there exist a distribution*

$$F(\mathbf{i}, \xi_\alpha, \alpha \in \chi) \tag{2.4}$$

and an $\mathbb{R}^2$ valued function $(y', y'') = \psi(\mathbf{i}, \xi_\alpha, \alpha \in \chi)$ such that when $\mathbf{I}$ and $\{\Xi_\alpha, \alpha \in \chi\}$ have distribution (2.4) then

$$(Y', Y'') = \psi(\mathbf{I}, \Xi_\alpha, \alpha \in \chi)$$

has distribution $F(y', y'')$. If $\mathbf{I}^\dagger$, $\{\Xi_\alpha^\dagger, \alpha \in \chi\}$ have distribution

$$dF^\dagger(\mathbf{i}, \xi_\alpha, \alpha \in \chi) = \frac{(y' - y'')^2}{\mathbb{E}(Y' - Y'')^2} dF(\mathbf{i}, \xi_\alpha, \alpha \in \chi) \tag{2.5}$$

then the pair

$$(Y^\dagger, Y^\ddagger) = \psi(\mathbf{I}^\dagger, \Xi_\alpha^\dagger, \alpha \in \chi)$$

has distribution $F^\dagger(y^\dagger, y^\ddagger)$ satisfying

$$dF^\dagger(y', y'') = \frac{(y' - y'')^2}{\mathbb{E}(Y' - Y'')^2} dF(y', y''). \tag{2.6}$$

In the following, for functions $f : \mathbb{R} \to \mathbb{R}$, we let $|f|_\infty = \sup_{x \in \mathbb{R}} |f(x)|$ be the supremum norm. Theorems 2.4 and 2.5 below provide respectively L^1 and L^∞ bounds between Y' and a standard normal random variable in term of Y^* satisfying 2.3.

Theorem 2.4. *Let Y' be a mean zero, variance $\sigma^2 > 0$ random variable, and Y^* and $R(Y')$ be defined on the same space as Y', satisfying (2.3), then*

$$d_1(\mathcal{L}(Y'/\sigma), \mathcal{L}(Z)) \leq \frac{2}{\sigma}\mathbb{E}|Y^* - Y'| + \sqrt{\frac{2}{\pi}}\frac{|\mathbb{E}Y'R|}{\sigma^2\lambda} + \frac{2\mathbb{E}|R|}{\sigma\lambda}, \tag{2.7}$$

where Z is a standard normal random variable.

Proof: For given $h \in \mathbb{L}$ let f be the unique bounded solution to the Stein equation

$$f'(w) - wf(w) = h(w) - Nh.$$

Then, (see e.g. Chen et al., 2011, Lemma 2.4),

$$|f|_\infty \leq 2, |f'|_\infty \leq \sqrt{\frac{2}{\pi}} \quad \text{and} \quad |f''|_\infty \leq 2.$$

Letting $g(x) = f(x/\sigma)$, we have $g'(x) = (1/\sigma)f'(x/\sigma)$, and thus

$$
\begin{aligned}
|\mathbb{E}h(Y'/\sigma) - Nh| &= \left| \mathbb{E}\left(f'\left(\frac{Y'}{\sigma}\right) - \frac{Y'}{\sigma}f\left(\frac{Y'}{\sigma}\right) \right) \right| \\
&= \left| \mathbb{E}\left(f'\left(\frac{Y'}{\sigma}\right) - \frac{1}{\sigma}Y'g\left(Y'\right) \right) \right| \\
&= \left| \mathbb{E}\left(f'\left(\frac{Y'}{\sigma}\right) - \sigma g'(Y^*) \right) + \frac{\mathbb{E}Y'R}{\sigma\lambda}\mathbb{E}g'(Y^*) - \frac{\mathbb{E}Rg(Y')}{\sigma\lambda} \right| \\
&\leq \mathbb{E}\left| f'\left(\frac{Y'}{\sigma}\right) - f'\left(\frac{Y^*}{\sigma}\right) \right| + \left| \frac{\mathbb{E}Y'R}{\sigma^2\lambda}\mathbb{E}f'\left(\frac{Y^*}{\sigma}\right) \right| + \left| \frac{\mathbb{E}Rf(Y'/\sigma)}{\sigma\lambda} \right| \\
&\leq \frac{2}{\sigma}\mathbb{E}|Y^* - Y'| + \sqrt{\frac{2}{\pi}}\frac{|\mathbb{E}Y'R|}{\sigma^2\lambda} + \frac{2\mathbb{E}|R|}{\sigma\lambda},
\end{aligned}
$$

where we have applied (2.3) with f replaced by g in the third equality.

$\square$

Theorem 2.5. *Let Y' be a mean zero, variance $\sigma^2 > 0$ random variable, and Y^* and $R(Y')$ be defined on the same space as Y', satisfying (2.3) and $|Y^* - Y| \leq \delta$. Then*

$$
d_\infty(\mathcal{L}(Y'/\sigma), \mathcal{L}(Z)) \leq \frac{\delta(1 + 1/\sqrt{2\pi} + \sqrt{2\pi}/4)}{\sigma} + \frac{|\mathbb{E}Y'R|}{\sigma^2\lambda} + \frac{\sqrt{2\pi}\mathbb{E}|R|}{4\sigma\lambda}, \qquad (2.8)
$$

where Z is a standard normal random variable.

Proof: We follow the proof of Theorem 5.1 of Chen et al. (2011) which obtained the the same type of bound using zero biasing. Let $z \in \mathbb{R}$, $\epsilon = \delta/\sigma$ and f be the solution of the equation

$$
f'(w) - wf(w) = \mathbf{1}_{\{w \leq z-\epsilon\}} - P(Z \leq z - \epsilon).
$$

Then, (see e.g. Chen et al., 2011, Lemma 2.3),

$$
|f|_\infty \leq \frac{\sqrt{2\pi}}{4}, |f'|_\infty \leq 1, \qquad (2.9)
$$

and for all w, u and v,

$$
|(w+u)f(w+u) - (w+v)f(w+v)| \leq (|w| + \sqrt{2\pi}/4)(|u| + |v|). \qquad (2.10)
$$

Then we have

$$
\begin{aligned}
f'(Y^*/\sigma) - (Y^*/\sigma)f(Y^*/\sigma) &= \mathbf{1}_{\{Y^*/\sigma \leq z-\epsilon\}} - P(Z \leq z - \epsilon) \\
&\leq \mathbf{1}_{\{Y'/\sigma \leq z\}} - P(Z \leq z - \epsilon).
\end{aligned}
$$

Letting $g(x) = f(x/\sigma)$, we have $g'(x) = (1/\sigma)f'(x/\sigma)$, and

$$
\begin{aligned}
P(Y'/\sigma \leq z) - P(Z < z) &= (P(Z < z - \epsilon) - P(Z < z)) \\
&\quad + P(Y'/\sigma \leq z) - P(Z < z - \epsilon) \\
&\geq -\epsilon/\sqrt{2\pi} + P(Y'/\sigma \leq z) - P(Z < z - \epsilon) \\
&\geq -\epsilon/\sqrt{2\pi} + \mathbb{E}[f'(Y^*/\sigma) - (Y^*/\sigma)f(Y^*/\sigma)] \\
&= -\epsilon/\sqrt{2\pi} + (1/\sigma)\mathbb{E}[\sigma^2 g'(Y^*) - Y^*g(Y^*)].
\end{aligned}
$$

Then, using (2.3) and applying (2.9) in the last inequality,

$$
\begin{aligned}
P(Y'/\sigma \le z) - P\left(Z < z\right) \;\ge\; & -\frac{\epsilon}{\sqrt{2\pi}} + \frac{1}{\sigma}\mathbb{E}\left[Y'g(Y') - Y^*g(Y^*)\right] \\
& + \frac{\mathbb{E}Y'R}{\sigma\lambda}\mathbb{E}g'(Y^*) - \frac{\mathbb{E}Rg(Y')}{\sigma\lambda} \\
=\; & -\frac{\epsilon}{\sqrt{2\pi}} + \mathbb{E}\left[\frac{Y'}{\sigma}f\left(\frac{Y'}{\sigma}\right) - \frac{Y^*}{\sigma}f\left(\frac{Y^*}{\sigma}\right)\right] \\
& + \frac{\mathbb{E}Y'R}{\sigma^2\lambda}\mathbb{E}f'\left(\frac{Y^*}{\sigma}\right) - \frac{\mathbb{E}Rf(Y'/\sigma)}{\sigma\lambda} \\
\ge\; & -\frac{\epsilon}{\sqrt{2\pi}} + \mathbb{E}\left[\frac{Y'}{\sigma}f\left(\frac{Y'}{\sigma}\right) - \frac{Y^*}{\sigma}f\left(\frac{Y^*}{\sigma}\right)\right] \\
& - \frac{|\mathbb{E}Y'R|}{\sigma^2\lambda} - \frac{\sqrt{2\pi}\mathbb{E}|R|}{4\sigma\lambda}. \qquad (2.11)
\end{aligned}
$$

Writing $\Delta = Y^*/\sigma - Y'/\sigma$ and applying (2.10) and that $|\Delta| \le \epsilon$ yields

$$
\begin{aligned}
\left|\mathbb{E}\left[\frac{Y'}{\sigma}f\left(\frac{Y'}{\sigma}\right) - \frac{Y^*}{\sigma}f\left(\frac{Y^*}{\sigma}\right)\right]\right| &= \left|\mathbb{E}\left[\frac{Y'}{\sigma}f\left(\frac{Y'}{\sigma}\right) - \left(\frac{Y'}{\sigma} + \Delta\right)f\left(\frac{Y'}{\sigma} + \Delta\right)\right]\right| \\
&\le \mathbb{E}\left[\left(\frac{|Y'|}{\sigma} + \frac{\sqrt{2\pi}}{4}\right)|\Delta|\right] \\
&\le \epsilon(1 + \sqrt{2\pi}/4) = \frac{\delta}{\sigma}(1 + \sqrt{2\pi}/4).
\end{aligned}
$$

Using this inequality in (2.11) yields

$$
P(Y'/\sigma \le z) - P\left(Z < z\right) \ge -\frac{\delta}{\sigma}\left(\frac{1}{\sqrt{2\pi}} + 1 + \frac{\sqrt{2\pi}}{4}\right) - \frac{|\mathbb{E}Y'R|}{\sigma^2\lambda} - \frac{\sqrt{2\pi}\mathbb{E}|R|}{4\sigma\lambda}.
$$

A similar argument yields the reverse inequality.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

We end up this section by mentioning a connection between our results in this section and the known ones when $R = 0$.

Remark 2.6. When $R = 0$, the construction of Lemma 2.1 yields the construction of the zero bias distribution as in Lemma 1.1. Furthermore, the L^1 and L^∞ bounds in Theorems 2.4 and 2.5 reduce to the bounds in (1.7) and (1.8), respectively.

3. Ewens measure

In this section, we briefly describe the Ewens measure and state some necessary properties. Let $\mathcal{S}_n$ denotes the symmetric group. The Ewens distribution $\mathcal{E}_\theta$ on the symmetric group $\mathcal{S}_n$ with parameter $\theta > 0$, was first introduced in Ewens (1972) and used in population genetics to describe the probabilities associated with the number of times that different alleles are observed in the sample; see also Arratia et al. (2003) for the description in mathematical context. In the following, we let $\mathbb{N}_k = [k, \infty) \cap \mathbb{Z}$ and for $x \in \mathbb{R}$, $n \in \mathbb{N}_1$, we use the notations

$$
x^{(n)} = x(x+1)\cdots(x+n-1) \quad \text{and} \quad x_{(n)} = x(x-1)\cdots(x-n+1).
$$

Given a permutation $\pi \in \mathcal{S}_n$, the Ewens measure is given by

$$P_\theta(\pi) = \frac{\theta^{\#(\pi)}}{\theta^{(n)}}, \tag{3.1}$$

where $\#(\pi)$ denotes the number of cycles of π. We note that $\mathcal{E}_\theta$ specializes to the uniform distribution over all permutations when $\theta = 1$.

The Ewens measure $\mathcal{E}_\theta$ can be defined equivalently in term of $c_1(\pi), \ldots, c_n(\pi)$ as follows,

$$P_\theta(c_1, \ldots, c_n) = \mathbf{1}\left\{\sum_{j=1}^n jc_j = n\right\} \frac{n!}{\theta^{(n)}} \prod_{j=1}^n \frac{\theta^{c_j}}{j^{c_j} c_j!}, \tag{3.2}$$

where $c_q(\pi)$ is the number of q cycles of π and we write c_q for $c_q(\pi)$ for simplicity.

A permutation $\pi_n \in \mathcal{S}_n$ with the distribution $\mathcal{E}_\theta$ can be constructed by the 'so called' the Chinese restaurant process (see e.g. Aldous, 1985 and Pitman, 1996), as follows. For $n = 1$, π_1 is the unique permutation that maps 1 to 1 in $\mathcal{S}_1$. For $n \geq 2$, we construct π_n from π_{n-1} by either adding n as a fixed point with probability $\theta/(\theta + n - 1)$, or by inserting n uniformly into one of $n - 1$ locations inside a cycle of π_{n-1}, so each with probability $1/(\theta + n - 1)$.

For σ a permutation of the elements of $[n]$ and $B \subset [n]$, we will consider the reduced permutation $\sigma \setminus B$ of the elements $[n] \setminus B$ whose cycle representation is obtained by deleting all elements of B in the cycle representation of σ. For instance, if $n = 5$ and the cycle representation of σ is $(1)(2435)$ and $B = \{1, 2\}$ then $\sigma \setminus B$ has representation (354). Also, let σ_B be the permutation whose cycle structure is obtained by taking the cycle structure of σ and removing all cycles that contain any element of B^c. Here, for instance, σ_B has cycle structure (1). With $\#(\tau)$ denoting the number of cycles of the permutation τ, we easily see that $\#(\sigma) = \#(\sigma/B) + \#(\sigma_B)$, as any cycle of σ either contains, or does not contain, some element of B^c.

For $B \subset [n]$, Propositions 3.1 and 3.2 that follow respectively provide the joint unconditional probability that $\pi(i) = \xi_i, i \in B$, and conditional probability that $\pi(i) = \xi_i, i \in B^c$ given $\{\pi(i), i \in B\}$ under the Ewens distribution.

Proposition 3.1. *Let π be a permutation of $[n]$ with distribution $\mathcal{E}_\theta$, $\theta > 0$ and $B \subseteq [n]$. Then, for $\xi_i, i \in B$ distinct elements of $[n]$,*

$$P(\pi(i) = \xi_i, i \in B) = \frac{\theta^{\#(\pi_B)}}{(\theta + n - 1)_{(|B|)}}.$$

Proof: We prove this lemma by induction on the size of B. Since it is clear by (3.1) that the distribution of π depends only on the number of cycles, it is sufficient to prove the result for $B_m = \{n - m + 1, n - m + 2, \ldots, n\}$ and $m \in [n]$. For $B_1 = \{n\}$, by the starting configuration in the construction of π via the Chinese restaurant process described above, we immediately have

$$P_\theta(\pi(n) = k) = \begin{cases} \frac{1}{\theta + n - 1} & \text{if } k \neq n, \\ \frac{\theta}{\theta + n - 1} & \text{if } k = n \end{cases} = \frac{\theta^{\#(\pi_{B_1})}}{(\theta + n - 1)_{(|B_1|)}}.$$

Now we assume that the claim is true for B_{m-1}. To prove the result for B_m, we recall that the Chinese restaurant process either adds $n - m + 1$ as a fixed point with probability $\theta/(\theta + n - m)$, or inserts $n - m + 1$ uniformly into one of $n - m$

locations inside a cycle of π_{n-m}, so each with probability $1/(\theta + n - m)$. Hence, using the assumption that the result holds for B_{m-1}, we have

$$
\begin{aligned}
P_\theta(\pi(i) = \xi_i, i \in B_m) &= \begin{cases} \frac{1}{\theta+n-m} \frac{\theta^{\#(\pi_{B_{m-1}})}}{(\theta+n-1)_{(m-1)}} & \text{if } n - m + 1 \text{ is not a fixed point} \\ \frac{\theta}{\theta+n-m} \frac{\theta^{\#(\pi_{B_{m-1}})}}{(\theta+n-1)_{(m-1)}} & \text{if } n - m + 1 \text{ is a fixed point} \end{cases} \\
&= \frac{\theta^{\#(\pi_{B_m})}}{(\theta + n - 1)_{(m)}} = \frac{\theta^{\#(\pi_{B_m})}}{(\theta + n - 1)_{(|B_m|)}}.
\end{aligned}
$$

$\square$

Proposition 3.2. *Let π be a permutation of $[n]$ with distribution $\mathcal{E}_\theta$, $\theta > 0$ and $B \subset [n]$. Then, for $\xi_i, i \in [n]$ distinct elements of $[n]$,*

$$
P(\pi(i) = \xi_i, i \in B^c | \pi(k) = \xi_k, k \in B) = \frac{\theta^{\#(\pi \backslash B)}}{\theta^{(|B^c|)}}.
$$

Proof: Using the definition of conditional probability and (3.1) and applying Proposition 3.1, we have

$$
\begin{aligned}
P(\pi(i) = \xi_i, i \in B^c | \pi(k) = \xi_k, k \in B) &= \frac{P(\pi(i) = \xi_i, i \in [n])}{P(\pi(k) = \xi_k, k \in B)} \\
&= \frac{\theta^{\#(\pi)}/\theta^{(n)}}{\theta^{\#(\pi_B)}/(\theta + n - 1)_{(|B|)}} \\
&= \frac{\theta^{\#(\pi \backslash B)}}{\theta^{(|B^c|)}},
\end{aligned}
$$

where we have used $\#(\pi \backslash B) = \#(\pi) - \#(\pi_B)$ in the last equality.

$\square$

The joint moments of $c = (c_1, \ldots, c_n)$ in the uniform case were established in Watterson (1974) (See also Arratia et al., 2003). Using the similar argument, in the following proposition, we generalize the result to the joint moments of $c = (c_1, \ldots, c_n)$ under the distribution $\mathcal{E}_\theta$ for any $\theta > 0$. Note that, with $\theta = 1$, the proposition below is exactly the same as the result in Watterson (1974) and Arratia et al. (2003).

Proposition 3.3. *Let $c = (c_1, \ldots, c_n)$ be a cycle type of $\pi \in \mathcal{S}_n$ with distribution $\mathcal{E}_\theta$ with $\theta > 0$. Then, for $m_1, \ldots, m_n \in \mathbb{N}_0$ with $m = \sum_{j=1}^n jm_j$,*

$$
\mathbb{E}\left(\prod_{j=1}^n (c_j)_{(m_j)}\right) = \left(\frac{n_{(m)}}{(\theta + n - 1)_{(m)}} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{m_j}\right) \mathbf{1}\left\{\sum_{j=1}^n jc_j \leq n\right\}.
$$

Proof: Using $(c_j)_{(m_j)}/c_j! = 1/(c_j - m_j)!$ when $c_j \geq m_j$, we have

$$
\mathbb{E}\left(\prod_{j=1}^n (c_j)_{(m_j)}\right) = \sum_{c \in \mathbb{N}_0^n} P_\theta(c) \prod_{j=1}^n (c_j)_{(m_j)}
$$

$$
= \sum_{c:c_j \geq m_j} \mathbf{1}\left\{\sum_{j=1}^n jc_j = n\right\} \frac{n!}{\theta^{(n)}} \prod_{j=1}^n \frac{(c_j)_{(m_j)}\theta^{c_j}}{j^{c_j} c_j!}
$$

$$
= \frac{n_{(m)}}{(\theta+n-1)_{(m)}} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{m_j} \sum_{d \in \mathbb{N}_0^n} \mathbf{1}\left\{\sum_{j=1}^n jd_j = n-m\right\} \frac{(n-m)!}{\theta^{(n-m)}} \prod_{j=1}^n \frac{\theta^{d_j}}{j^{d_j} d_j!}
$$

$$
= \frac{n_{(m)}}{(\theta+n-1)_{(m)}} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{m_j} \sum_{d \in \mathbb{N}_0^n} \mathbf{1}\left\{\sum_{j=1}^{n-m} jd_j = n-m\right\} \frac{(n-m)!}{\theta^{(n-m)}} \prod_{j=1}^{n-m} \frac{\theta^{d_j}}{j^{d_j} d_j!}
$$

$$
= \left(\frac{n_{(m)}}{(\theta+n-1)_{(m)}} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{m_j}\right) \mathbf{1}\left\{\sum_{j=1}^n jc_j \leq n\right\}
$$

where d_j corresponds to $c_j - m_j$ and the sum over d in the second last line is one as it is taken over all possibilities of $\mathcal{S}_{n-m}$.

$\square$

4. Combinatorial CLT under the Ewens measure

In this section, following Section 6.1 of Chen et al. (2011), we study the distribution of

$$
Y = \sum_{i=1}^n a_{i,\pi(i)} \tag{4.1}
$$

where $A \in \mathbb{R}^{n \times n}$ is a given real matrix with components $\{a_{i,j}\}_{i,j=1}^n$ and $\pi \in \mathcal{S}_n$ has the Ewens distribution.

A distribution on $\mathcal{S}_n$ is said to be constant on cycle type if the probability of any permutation $\pi \in \mathcal{S}_n$ depends only on the cycle type $(c_1, \ldots, c_n)$. The work Chen et al. (2011) studied the distribution of (4.1) where $\pi \in \mathcal{S}_n$ has distribution constant on cycle type with no fixed points. It follows from (3.2) directly that the Ewens distribution is constant on cycle type and allows fixed point. Therefore, though several techniques in Section 6.1.2 of Chen et al. (2011) apply here, the main proofs and the coupling construction do not.

Letting

$$
a_{\bullet,\bullet} = \frac{1}{n(\theta+n-1)} \left(\theta \sum_{i=1}^n a_{i,i} + \sum_{i \neq j} a_{ij}\right), \tag{4.2}
$$

applying Proposition 3.1 with $B = \{i\}$, we have

$$
\mathbb{E}Y = \sum_{i=1}^n \mathbb{E}a_{i,\pi(i)} = \sum_{i=1}^n \left(\frac{\theta}{\theta+n-1} a_{i,i} + \sum_{j,j \neq i} \frac{1}{\theta+n-1} a_{i,j}\right) = na_{\bullet,\bullet}. \tag{4.3}
$$

Letting

$$\widehat{a}_{i,j} = a_{i,j} - a_{\bullet,\bullet},$$

and using (4.3), we have

$$\mathbb{E}\left[\sum_{i=1}^{n}\widehat{a}_{i,\pi(i)}\right] = \mathbb{E}\left[\sum_{i=1}^{n}(a_{i,\pi(i)} - a_{\bullet,\bullet})\right] = \mathbb{E}\left[\sum_{i=1}^{n}a_{i,\pi(i)} - na_{\bullet,\bullet}\right] = 0.$$

As a consequence, replacing $a_{i,j}$ by $\widehat{a}_{i,j}$, we may without loss of generality assume that

$$\mathbb{E}Y = 0 \quad \text{and} \quad a_{\bullet,\bullet} = 0, \tag{4.4}$$

and for simplicity, as in Chen et al. (2011), we consider only the symmetric case, that is, $a_{i,j} = a_{j,i}$ for all $i, j \in [n]$.

To rule out trivial cases, we assume in what follows that $\sigma^2 > 0$. We later calculate σ^2 explicitly in (4.23) of Lemma 4.8 and discuss in Remark 4.9 that it is of order n when the elements of A are well chosen in some sense. In this case, there exists $N \in \mathbb{N}_1$ such that $\sigma^2 > 0$ for $n > N$.

In the following theorem, we obtain upper bounds for the L^1 and L^∞ distances between Y given in (4.1) with the distribution $\mathcal{E}_\theta$ and a standard normal random variable and lower bounds for the L^∞ distance in the special case that the matrix A is integer-valued. Below, we consider the symbols π and Y interchageable with π' and Y', respectively.

Theorem 4.1. *Let $n \geq 6$ and $\{a_{i,j}\}_{i,j=1}^{n}$ be an array of real numbers satisfying*

$$a_{i,j} = a_{j,i}.$$

Let $\pi' \in \mathcal{S}_n$ be a random permutation with the distribution $\mathcal{E}_\theta$, with $\theta > 0$. Then, with Y' the sum in (4.1) with π' replacing π, assuming $\mathrm{Var}(Y') = \sigma^2 > 0$, and letting $W = (Y' - \mathbb{E}Y')/\sigma$ and Z a standard normal random variable,

$$d_1(\mathcal{L}(W), \mathcal{L}(Z)) \leq \frac{\alpha_1(\theta, M, n)}{\sigma}, \tag{4.5}$$

and

$$d_\infty(\mathcal{L}(W), \mathcal{L}(Z)) \leq \frac{\alpha_2(\theta, M, n)}{\sigma}, \tag{4.6}$$

where

$$\begin{aligned}
\alpha_1(\theta, M, n) &= 40M + \kappa_{\theta,n,1}\sqrt{2/\pi}M\left(3 + \frac{\theta+1}{n-1}\right) + \frac{\kappa_{\theta,n,2}\sqrt{2/\pi}M}{n-1} \\
&\quad + \theta M\left(1.2\sqrt{2/\pi} + \frac{1.2(6n+4\theta-5)}{\theta+n-1} + \frac{\theta n}{(\theta+n-1)_{(2)}}\right),
\end{aligned}$$

and

$$\begin{aligned}
\alpha_2(\theta, M, n) &= 20(1 + 1/\sqrt{2\pi} + \sqrt{2\pi}/4)M + \kappa_{\theta,n,1}M\left(3 + \frac{\theta+1}{n-1}\right) + \frac{\kappa_{\theta,n,2}M}{n-1} \\
&\quad + \theta M\left(1.2 + \frac{0.15\sqrt{2\pi}(6n+4\theta-5)}{\theta+n-1} + \frac{0.125\sqrt{2\pi}\theta n}{(\theta+n-1)_{(2)}}\right)
\end{aligned}$$

with

$$M = \max_{i,j}|a_{i,j} - a_{\bullet,\bullet}|, \; a_{\bullet,\bullet} \; \text{ as in (4.2)},$$

$$\kappa_{\theta,n,1} = \sqrt{\frac{\theta^2 n_{(2)}}{(\theta + n - 1)_{(2)}} + \frac{\theta n}{\theta + n - 1}} \tag{4.7}$$

and

$$\kappa_{\theta,n,2} = \sqrt{\frac{\theta^4 n_{(4)}}{(\theta + n - 1)_{(4)}} + \frac{4\theta^3 n_{(3)}}{(\theta + n - 1)_{(3)}} + \frac{2\theta^2 n_{(2)}}{(\theta + n - 1)_{(2)}}}. \tag{4.8}$$

In particular, if $a_{i,j}$, $i, j \in [n]$ are all integers, then

$$\frac{1}{6\sqrt{3}\sigma + 3} \leq d_\infty(\mathcal{L}(W), \mathcal{L}(Z)) \leq \frac{\alpha_2(\theta, M, n)}{\sigma}. \tag{4.9}$$

Remark 4.2 below discusses the behavior of (4.7) and (4.8) in θ and the bounds on the rates of convergence in n.

Remark 4.2.   (1) $\kappa_{\theta,n,1}$ and $\kappa_{\theta,n,2}$ given in (4.7) and (4.8) are of constant order in n since

$$\kappa_{\theta,n,1} \to \sqrt{\theta^2 + \theta} \quad \text{and} \quad \kappa_{\theta,n,2} \to \sqrt{\theta^4 + 4\theta^3 + 2\theta^2} \quad \text{as} \quad n \to \infty.$$

In particular, $\kappa_{1,n,1} = \sqrt{2}$ and $\kappa_{1,n,2} = \sqrt{7}$.
(2) Since $\kappa_{\theta,n,1}$ and $\kappa_{\theta,n,2}$ are of constant order in n, $\alpha_1(\theta, M, n)$ and $\alpha_2(\theta, M, n)$ are also of constant order in n for all $\theta > 0$. Therefore, the L^1 and L^∞ bounds on the right hand side of (4.5) and (4.6) are of order σ^{-1} which is of the same order as the L^1 bound in the uniform case in Theorem 4.8 of Chen et al. (2011). The uniform distribution corresponds to the special case of the Ewens distribution with $\theta = 1$ and the result in this case is presented in Corollary 4.3 below. The order of σ^2 in n is considered in Remark 4.9, where we find that if $A = [a_{i,j}]$ is chosen well in some sense then σ^2 will be of order n, implying that the bounds in (4.5) and (4.6) are of order $n^{-1/2}$.
(3) In the case that $a_{i,j}$, $i, j \in [n]$ are interger-valued, the L^∞ upper and lower bounds in (4.9) are of the same order σ^{-1}, which is thus the optimal order for the L^∞ distance.
(4) It is easy to see that $\alpha_1(\theta, M, n)$ and $\alpha_2(\theta, M, n)$ are increasing in θ and thus so are the L^1 and L^∞ bounds in (4.5) and (4.6). In fact, the expressions $\alpha_1(\theta, M, n)$ and $\alpha_2(\theta, M, n)$ are obtained from (2.7) and (2.8), respectively, which depend on R. This remainder R, given explicitly in Lemma 4.6 below, depends on the number of fixed points of π, which become more likely as θ increases.
(5) As θ tends to zero,

$$\alpha_1(\theta, M, n) \to 40M \quad \text{and} \quad \alpha_2(\theta, M, n) \to 20(1 + 1/\sqrt{2\pi} + \sqrt{2\pi}/4)M.$$

Therefore, the L^1 and L^∞ bounds on the right hand side of (4.5) and (4.6) converge to $40M/\sigma$ and $20(1 + 1/\sqrt{2\pi} + \sqrt{2\pi}/4)M/\sigma$, respectively. This corresponds to the case where a large number of cycles is unlikely.

Next we present Corollary 4.3 that specializes Theorem 4.1 to the case where the random permutation π' has the uniform distribution, corresponding to the special case of the Ewens distribution with $\theta = 1$. Indeed, the result immediately follows from Theorem 4.1 by applying the bounds

$$\alpha_1(1, M, n) = \left(47.2 + 6/\sqrt{\pi} + 1.2\sqrt{2/\pi} + \frac{5 + \sqrt{14}}{\sqrt{\pi}(n - 1)} - \frac{1.2}{n}\right) M \leq 53M$$

and

$$\alpha_2(1, M, n) = \left(21.2 + 3\sqrt{2} + 5.9\sqrt{2\pi} + 20/\sqrt{2\pi} \right.$$

$$\left. + \frac{2\sqrt{2} + \sqrt{7} + 0.125\sqrt{2\pi}}{n-1} - \frac{0.15\sqrt{2\pi}}{n} \right) M \leq 50M,$$

which hold for all $n \geq 6$.

Corollary 4.3. *Let $n \geq 6$ and $\{a_{i,j}\}_{i,j=1}^n$ be an array of real numbers satisfying*

$$a_{i,j} = a_{j,i}.$$

Let π' be a random permutation with the uniform distribution over $\mathcal{S}_n$. Then, with Y' the sum in (4.1) with π' replacing π, assuming $\mathrm{Var}(Y') = \sigma^2 > 0$, and letting $W = (Y' - \mathbb{E}Y')/\sigma$ and Z a standard normal random variable,

$$d_1(\mathcal{L}(W), \mathcal{L}(Z)) \leq \frac{53M}{\sigma},$$

and

$$d_\infty(\mathcal{L}(W), \mathcal{L}(Z)) \leq \frac{50M}{\sigma},$$

where

$$M = \max_{i,j} |a_{i,j} - a_{\bullet,\bullet}| \quad and \quad a_{\bullet,\bullet} \quad is \ as \ in \ (4.2).$$

In particular, if $a_{i,j}$, $i, j \in [n]$ are all integers, then

$$\frac{1}{6\sqrt{3}\sigma + 3} \leq d_\infty(\mathcal{L}(W), \mathcal{L}(Z)) \leq \frac{50M}{\sigma}.$$

Section 6.1.2 of Chen et al. (2011) proved the main L^1 and L^∞ bounds by first considering each cycle type $c := (c_1, \ldots, c_n)$ separately and then combining all cases. Since fixed points are allowed in the present work, $P_\theta(\pi(i) = i|c)$ varies as c changes. This dependence on c makes it difficult to follow the same proof structure and coupling construction as there. As a result, we construct a new coupling for the proof of Theorem 4.1.

To prove the main results of this section, our target is therefore to apply Theorems 2.4 and 2.5. Hence we will first construct an approximate λ, R-Stein pair and then an approximate zero bias coupling through the helps of Lemmas 2.1 and 2.3. For this purpose, we present a sequence of lemmas below. The first two lemmas were proved in Chen et al. (2011). The proofs of Lemmas 4.6 and 4.8, though important, the first closely follows the proof of Lemma 6.9 of Chen et al. (2011) and the latter is straightforward, can be found in the Appendix. Lemma 4.4 forms a partition of the space based on the cycle structure of $\pi \in \mathcal{S}_n$. Using that partition, Lemma 4.5 expresses the difference in the values taken on by the exchangeable pair coupling given in Lemma 4.6. Below, for $i, j \in [n]$, we write $i \sim j$ if i and j are in the same cycle, let $|i|$ be the length of the cycle containing i and let $\tau_{i,j}$, $i, j \in [n]$ be the permutation that transposes i and j.

Lemma 4.4 (Chen et al., 2011)**.** *Let π be a fixed permutation. For any $i \neq j$, distinct elements of $[n]$, the sets $A_0, \ldots, A_5$ form a partition of the space where,*

$$A_0 = \{|\{i, j, \pi(i), \pi(j)\}| = 2\}$$
$$A_1 = \{|i| = 1, |j| \geq 2\}, \quad A_2 = \{|i| \geq 2, |j| = 1\}$$
$$A_3 = \{|i| \geq 3, \pi(i) = j\}, \quad A_4 = \{|j| \geq 3, \pi(j) = i\} \quad \text{and}$$
$$A_5 = \{|\{i, j, \pi(i), \pi(j)\}| = 4\}.$$

Additionally, the sets $A_{0,1}$ and $A_{0,2}$ partition A_0 where

$$A_{0,1} = \{\pi(i) = i, \pi(j) = j\}, \quad A_{0,2} = \{\pi(i) = j, \pi(j) = i\},$$

and we may also write

$$A_1 = \{\pi(i) = i, \pi(j) \neq j\}, \quad A_2 = \{\pi(i) \neq i, \pi(j) = j\}$$
$$A_3 = \{\pi(i) = j, \pi(j) \neq i\}, \quad A_4 = \{\pi(j) = i, \pi(i) \neq j\},$$

and membership in A_m, $m = 0, \ldots, 4$ depends only on $i, j, \pi(i), \pi(j)$.
  Lastly, the sets $A_{5,m}$, $m = 1, \ldots, 4$ partition A_5, where

$$A_{5,1} = \{|i| = 2, |j| = 2, i \not\sim j\}$$
$$A_{5,2} = \{|i| = 2, |j| \geq 3\}, \quad A_{5,3} = \{|i| \geq 3, |j| = 2\}$$
$$A_{5,4} = \{|i| \geq 3, |j| \geq 3\} \cap A_5,$$

and membership in $A_{5,m}$, $m = 1, \ldots, 4$ depends only on $i, j, \pi^{-1}(i), \pi^{-1}(j), \pi(i),$ *$\pi(j)$.*

  We now state Lemma 6.8 of Chen et al. (2011), noting that y' and y'' there are incorrectly interchanged on the left hand side of (4.10).

Lemma 4.5 (Chen et al., 2011)**.** *Let π be a fixed permutation and $i \neq j$ distinct elements of $[n]$. Letting $\pi(-\alpha) = \pi^{-1}(\alpha)$ for $\alpha \in [n]$ set*

$$\chi_{i,j} = \{-i, -j, i, j\}, \text{ so that } \{\pi(\alpha), \alpha \in \chi_{i,j}\} = \{\pi^{-1}(i), \pi^{-1}(j), \pi(i), \pi(j)\}.$$

Then, for $\pi'' = \tau_{i,j}\pi'\tau_{i,j}$, and y' and y'' given by (4.1) with π' and π'' replacing π, respectively,

$$y' - y'' = b(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) \tag{4.10}$$

where

$$b(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) = \sum_{m=0}^{5} b_m(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) \mathbf{1}_{A_m} \tag{4.11}$$

with A_m, $m = 0, \ldots, 5$ as in Lemma 4.4, $b_0(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) = 0$,

$$b_1(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) = a_{i,i} + a_{\pi^{-1}(j),j} + a_{j,\pi(j)} - (a_{j,j} + a_{\pi^{-1}(j),i} + a_{i,\pi(j)}),$$
$$b_2(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) = a_{j,j} + a_{\pi^{-1}(i),i} + a_{i,\pi(i)} - (a_{i,i} + a_{\pi^{-1}(i),j} + a_{j,\pi(i)}),$$
$$b_3(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) = a_{\pi^{-1}(i),i} + a_{i,j} + a_{j,\pi(j)} - (a_{\pi^{-1}(i),j} + a_{j,i} + a_{i,\pi(j)}),$$
$$b_4(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) = a_{\pi^{-1}(j),j} + a_{j,i} + a_{i,\pi(i)} - (a_{\pi^{-1}(j),i} + a_{i,j} + a_{j,\pi(i)}),$$
and

$$b_5(i, j, \pi(\alpha), \alpha \in \chi_{i,j}) = a_{\pi^{-1}(i),i} + a_{i,\pi(i)} + a_{\pi^{-1}(j),j} + a_{j,\pi(j)}$$
$$- (a_{\pi^{-1}(i),j} + a_{j,\pi(i)} + a_{\pi^{-1}(j),i} + a_{i,\pi(j)}). \tag{4.12}$$

Next we construct an exchangeable pair satisfying (1.9), proved in the Appendix, which we call an approximate λ, R-Stein pair.

Lemma 4.6. *For $n \geq 6$, let $\{a_{i,j}\}_{i,j=1}^n$ be an array of real numbers satisfying $a_{i,j} = a_{j,i}$ and $a_{\bullet,\bullet} = 0$ where $a_{\bullet,\bullet}$ is as in (4.2). Let $\pi' \in \mathcal{S}_n$ a random permutation has the Ewens measure $\mathcal{E}_\theta$ with $\theta > 0$. Further, let I, J be chosen independently of π, uniformly from all pairs of distinct elements of $\{1, \ldots n\}$. Then, letting $\pi'' = \tau_{I,J}\pi'\tau_{I,J}$ and Y' and Y'' be as in (4.1) with π' and π'' replacing π, respectively, (Y', Y'') is an approximate $4/n, R$-Stein pair with*

$$R(Y') = \frac{1}{n(n-1)}\mathbb{E}[T|Y'] \tag{4.13}$$

where

$$T = 2(n + c_1 - 2(\theta+1)) \sum_{|i|=1} a_{i,i} + 2(c_1 - 2\theta) \sum_{|i|\geq 2} a_{i,i}$$
$$-4 \sum_{|i|,|j|=1, j\neq i} a_{i,j} - 4 \sum_{|i|=1, |j|\geq 2} a_{i,j}. \tag{4.14}$$

The next lemma provides bounds for $\mathbb{E}|R|$ and $|\mathbb{E}Y'R|$ that will be used when applying Theorems 2.4 and 2.5. To derive the bounds in Lemma 4.7, we use consequences of Proposition 3.3, which can be easily verified, that

$$\mathbb{E}[c_1] = \frac{\theta n}{\theta + n - 1}, \quad \mathbb{E}[c_1(c_1 - 1)] = \frac{\theta^2 n_{(2)}}{(\theta + n - 1)_{(2)}},$$

$$\mathbb{E}[c_1^2] = \frac{\theta^2 n_{(2)}}{(\theta + n - 1)_{(2)}} + \frac{\theta n}{\theta + n - 1}, \quad \text{and}$$

$$\mathbb{E}[c_1^2(c_1 - 1)^2] = \frac{\theta^4 n_{(4)}}{(\theta + n - 1)_{(4)}} + \frac{4\theta^3 n_{(3)}}{(\theta + n - 1)_{(3)}} + \frac{2\theta^2 n_{(2)}}{(\theta + n - 1)_{(2)}}. \tag{4.15}$$

Lemma 4.7. *Let (Y', Y'') be an approximate $4/n, R$-Stein pair constructed as in Lemma 4.6 with R as in (4.13). Then*

$$\mathbb{E}|R| \leq \frac{\theta M(12n + 8\theta - 10)}{(n-1)(\theta + n - 1)} + \frac{2\theta^2 M}{(\theta + n - 1)_{(2)}}, \tag{4.16}$$

and

$$|\mathbb{E}Y'R| \leq \left(10\kappa_{\theta,n,1} + 4\theta + \frac{4(\kappa_{\theta,n,1}(\theta+1) + \kappa_{\theta,n,2})}{n}\right)\frac{M\sigma}{n-1} \tag{4.17}$$

where $M = \max_{i,j}|a_{i,j} - a_{\bullet,\bullet}|$ with $a_{\bullet,\bullet}$ as in (4.2) and $\kappa_{\theta,n,1}$ and $\kappa_{\theta,n,2}$ are given in (4.7) and (4.8), respectively.

Proof: By replacing $a_{i,j}$ by $a_{i,j} - a_{\bullet,\bullet}$ we may assume (4.4) is satisfied, that is, that $\mathbb{E}Y' = 0$ and $a_{\bullet,\bullet} = 0$ and thus demonstrate the claim with $M = \max_{i,j}|a_{i,j}|$.

We start with the first claim, using conditional Jensen's inequality and (4.13) to obtain $|R| \leq \mathbb{E}[|T| \,|Y']/(n(n-1))$ and therefore that $E|R| \leq E|T|/(n(n-1))$,

where T is given by (4.14). Using $|a_{i,j}| \le M$ for all i, j, we have

$$\mathbb{E}|R| \quad \le \quad \frac{2M\mathbb{E}|(n + c_1 - 2(\theta + 1))c_1|}{n(n-1)} + \frac{2M\mathbb{E}|(c_1 - 2\theta)(n - c_1)|}{n(n-1)}$$
$$+ \frac{4M\mathbb{E}[c_1(c_1 - 1)]}{n(n-1)} + \frac{4M\mathbb{E}[c_1(n - c_1)]}{n(n-1)}$$
$$= \quad \frac{2M\mathbb{E}|(n-1)c_1 + c_1(c_1 - 1) - 2\theta c_1|}{n(n-1)}$$
$$+ \frac{2M\mathbb{E}|(c_1 - 2\theta)(n - c_1)|}{n(n-1)} + \frac{4M\mathbb{E}[(n-1)c_1]}{n(n-1)}$$
$$\le \quad \frac{2M\mathbb{E}[(n-1)c_1 + c_1(c_1 - 1) + 2\theta c_1]}{n(n-1)} + \frac{2M\mathbb{E}[nc_1 + 2\theta n]}{n(n-1)} + \frac{4M\mathbb{E}[c_1]}{n}.$$

Now using (4.15), we obtain

$$\mathbb{E}|R| \quad \le \quad \frac{2\theta M}{\theta + n - 1} + \frac{2\theta^2 M}{(\theta + n - 1)_{(2)}} + \frac{4\theta^2 M}{(n-1)(\theta + n - 1)}$$
$$+ \frac{2\theta Mn}{(n-1)(\theta + n - 1)} + \frac{4\theta M}{n - 1} + \frac{4\theta M}{\theta + n - 1}.$$

Combining similar terms in the last expression yields the bound in (4.16).

Now we consider the second claim. First note that by (4.15),

$$\mathbb{E}\left[\left(\sum_{|i|=1} a_{i,i}\right)^2\right] \le M^2\mathbb{E}[c_1^2] = \kappa_{\theta,n,1}^2 M^2.$$

Hence, using that

$$n(n-1)|E[Y'R(Y')]| = |E[Y'E[T|Y']]| = |E[E[Y'T|Y']]| = |E[Y'T]|,$$

for the first term in the product $Y'T$, bounding c_1 by n, we obtain

$$\left| \frac{\mathbb{E}\left(2(n + c_1 - 2(\theta + 1))Y' \sum_{|i|=1} a_{i,i}\right)}{n(n-1)} \right|$$
$$\le \left| \frac{4(n + \theta + 1)\mathbb{E}\left(Y' \sum_{|i|=1} a_{i,i}\right)}{n(n-1)} \right|$$
$$\le \frac{4\kappa_{\theta,n,1}M\sigma}{n - 1} + \frac{4(\theta + 1)\kappa_{\theta,n,1}M\sigma}{n(n-1)} \qquad (4.18)$$

where here, and at similar steps later on, we apply the Cauchy–Schwarz inequality.

Now, to bound the second term in the $Y'T$ product, by (4.15) we have

$$\mathbb{E}\left[c_1^2\left(\sum_{|i|\ge 2} a_{i,i}\right)^2\right] \le M^2 n^2 \mathbb{E}[c_1^2] = \kappa_{\theta,n,1}^2 M^2 n^2$$

and

$$\mathbb{E}\left[\left(\sum_{|i|\geq 2} a_{i,i}\right)^2\right] \leq M^2 n^2.$$

Thus, for that second term,

$$\left|\frac{\mathbb{E}\left((2c_1(\pi)-4\theta)Y'\sum_{|i|\geq 2} a_{i,i}\right)}{n(n-1)}\right|$$

$$\leq \left|\frac{2\mathbb{E}\left(c_1 Y'\sum_{|i|\geq 2} a_{i,i}\right)}{n(n-1)}\right| + \left|\frac{4\theta\mathbb{E}\left(Y'\sum_{|i|\geq 2} a_{i,i}\right)}{n(n-1)}\right|$$

$$\leq \frac{2\kappa_{\theta,n,1}M\sigma}{n-1} + \frac{4\theta M\sigma}{n-1}. \tag{4.19}$$

For the third term, again by (4.15),

$$\mathbb{E}\left[\left(\sum_{|i|=1,|j|=1,j\neq i} a_{i,j}\right)^2\right] \leq M^2\mathbb{E}[c_1^2(c_1-1)^2] = M^2\kappa_{\theta,n,2}^2,$$

and thus,

$$\left|\frac{4\mathbb{E}\left(Y'\sum_{|i|=1,|j|=1,j\neq i} a_{i,j}\right)}{n(n-1)}\right| \leq \frac{4\kappa_{\theta,n,2}M\sigma}{n(n-1)}. \tag{4.20}$$

For the fourth term, by (4.15),

$$\mathbb{E}\left[\left(\sum_{|i|=1,|j|\geq 2} a_{i,j}\right)^2\right] \leq n^2 M^2\mathbb{E}[c_1^2] = n^2 M^2\kappa_{\theta,n,1}^2,$$

and thus,

$$\left|\frac{4\mathbb{E}\left(Y'\sum_{|i|=1,|j|\geq 2} a_{i,j}\right)}{n(n-1)}\right| \leq \frac{4\kappa_{\theta,n,1}M\sigma}{n-1}. \tag{4.21}$$

Hence, summing (4.18) to (4.21), we have

$$\begin{aligned}
|\mathbb{E}Y'R| &\leq \frac{4\kappa_{\theta,n,1}M\sigma}{n-1} + \frac{4(\theta+1)\kappa_{\theta,n,1}M\sigma}{n(n-1)} + \frac{2\kappa_{\theta,n,1}M\sigma}{n-1} \\
&\quad + \frac{4\theta M\sigma}{n-1} + \frac{4\kappa_{\theta,n,2}M\sigma}{n(n-1)} + \frac{4\kappa_{\theta,n,1}M\sigma}{n-1}.
\end{aligned}$$

Combining similar terms and factoring $M\sigma/(n-1)$ out in the last expression yield the bound in (4.17) and thus completes the proof.

$\square$

Next, Lemma 4.8, proved in the Appendix, provides detailed expressions for $\mathbb{E}[(Y'-Y'')^2]$ and σ^2. The order of σ^2 will be discussed in Remark 4.9.

Lemma 4.8. *Let $n \geq 6$ and (Y', Y'') be an approximate $4/n, R$-Stein pair constructed as in Lemma 4.6 with R as in (4.13). Then*

$$\mathbb{E}[(Y' - Y'')^2] = 2\beta_1 + 2\beta_3 + \beta_{5,1} + 2\beta_{5,2} + \beta_{5,4}, \tag{4.22}$$

and

$$\sigma^2 = \frac{n}{8}(2\beta_1 + 2\beta_3 + \beta_{5,1} + 2\beta_{5,2} + \beta_{5,4}) + \frac{n\mathbb{E}Y'R}{4}, \tag{4.23}$$

where, with b_m, $m = 1, \ldots, 5$ are given in (4.12),

$$\beta_1(\theta, n) = \frac{1}{n(n-1)(\theta + n - 1)_{(3)}}$$

$$\times \left(\theta^2 \sum_{i,j,s} b_1^2(i, j, i, s, i, s) + \theta \sum_{i,j,s,l} b_1^2(i, j, i, s, i, l) \right), \tag{4.24}$$

$$\beta_3(\theta, n) = \frac{1}{n(n-1)(\theta + n - 1)_{(3)}}$$

$$\times \left(\theta \sum_{i,j,r} b_3^2(i, j, r, i, j, r) + \sum_{i,j,r,l} b_3^2(i, j, r, i, j, l) \right), \tag{4.25}$$

$$\beta_{5,1}(\theta, n) = \frac{\theta^2}{n(n-1)(\theta + n - 1)_{(4)}} \sum_{i,j,r,s} b_5^2(i, j, r, s, r, s), \tag{4.26}$$

$$\beta_{5,2}(\theta, n) = \frac{\theta}{n(n-1)(\theta + n - 1)_{(4)}} \sum_{i,j,r,s,l} b_5^2(i, j, r, s, r, l), \tag{4.27}$$

and

$$\beta_{5,4}(\theta, n) = \frac{1}{n(n-1)(\theta + n - 1)_{(4)}} \left(\theta \sum_{i,j,r,k} b_5^2(i, j, r, k, k, r) \right.$$

$$\left. + \sum_{i,j,r,k,l} b_5^2(i, j, r, k, k, l) + \sum_{i,j,r,s,k,l} b_5^2(i, j, r, s, k, l) \right). \tag{4.28}$$

Remark 4.9. Here we consider the order of σ^2 given in (4.23). Let β_1, β_3, $\beta_{5,1}$, $\beta_{5,2}$ and $\beta_{5,4}$ be given as in (4.24)-(4.28), respectively. Using (4.17), we have

$$\begin{aligned} \sigma^2 &= \frac{n}{8}(2\beta_1 + 2\beta_3 + \beta_{5,1} + 2\beta_{5,2} + \beta_{5,4}) + \frac{n\mathbb{E}Y'R}{4} \\ &\geq \frac{n}{8}(2\beta_1 + 2\beta_3 + \beta_{5,1} + 2\beta_{5,2} + \beta_{5,4}) - \frac{n|\mathbb{E}Y'R|}{4} \\ &\geq \frac{n}{8}(2\beta_1 + 2\beta_3 + \beta_{5,1} + 2\beta_{5,2} + \beta_{5,4}) \\ &\quad - \left(3\kappa_{\theta,n,1} + 1.2\theta + \frac{1.2(\kappa_{\theta,n,1}(\theta+1) + \kappa_{\theta,n,2})}{n} \right) M\sigma. \end{aligned}$$

As discussed in Remark 4.2, $\kappa_{\theta,n,1}$ and $\kappa_{\theta,n,2}$ are of constant order in n and thus σ^2 is of order n whenever at least one of $\beta_\gamma(\theta, n)$, $\gamma = 1, 3, (5, 1), (5, 2), (5, 4)$ are of constant order. Since the final sum of (4.28) in the expression for $\beta_{5,4}(\theta, n)$ has $n_{(6)}$ terms and the denominator is of order n^6, $\beta_{5,4}(\theta, n)$ is of constant order in n if

the elements of $\{a_{i,j} : \{i, j\} \subset [n]\}$ are chosen so that the values do not depend on n and $b_5(i, j, r, s, k, l)$ with distinct $i, j, r, s, k, l \in [n]$ are nonzero for at least $\lfloor \delta n^6 \rfloor$ terms for some $\delta > 0$. For instance, if $a_{i,j}$, $\{i, j\} \subset [n]$ are independent identically distributed uniform random variables on $[0, 1]$ then these sufficient conditions hold almost surely.

Now we are in the final step before proving Theorem 4.1, that is, to construct an approximate zero bias coupling that will be used when applying Theorems 2.4 and 2.5. Prior to doing that, we first specialize the outline in Lemma 2.3 to the more specific case where the random index $\mathbf{I}$ is chosen independently of the permutation and

$$Y'' - Y' = f(\mathbf{I}, \Xi_\alpha, \alpha \in \chi_{\mathbf{I}}), \tag{4.29}$$

where $\mathbf{I}$ and $\chi_{\mathbf{I}} \subset \chi$ are vectors of small dimensions and f is a function with range being subset of $\mathbb{R}$, that is, we consider situations that $Y'' - Y'$ depends on only a few variables.

Letting $Y', Y'', Y^\dagger, Y^\ddagger$ be constructed as in Lemma 2.3 satisfying (4.29), we follow Section 4.4.1 of Chen et al. (2011) decomposing $P(\mathbf{i}, \xi_\alpha, \alpha \in \chi)$ as

$$P(\mathbf{i}, \xi_\alpha, \alpha \in \chi) = P(\mathbf{I} = \mathbf{i})P_{\mathbf{i}}(\xi_\alpha, \alpha \in \chi_{\mathbf{i}})P_{\mathbf{i}^c|\mathbf{i}}(\xi_\alpha, \alpha \notin \chi_{\mathbf{i}}|\xi_\alpha, \alpha \in \chi_{\mathbf{i}}), \tag{4.30}$$

where $P_{\mathbf{i}}(\xi_\alpha, \alpha \in \chi_{\mathbf{i}})$ is the marginal distribution of ξ_α for $\alpha \in \chi_{\mathbf{i}}$, and $P_{\mathbf{i}^c|\mathbf{i}}(\xi_\alpha, \alpha \notin \chi_{\mathbf{i}}|\xi_\alpha, \alpha \in \chi_{\mathbf{i}})$ the conditional distribution of ξ_α for $\alpha \notin \chi_{\mathbf{i}}$ given ξ_α for $\alpha \in \chi_{\mathbf{i}}$.

For the square bias distribution, similarly, we decompose $P^\dagger(\mathbf{i}, \xi_\alpha, \alpha \in \chi)$ as

$$P^\dagger(\mathbf{i}, \xi_\alpha, \alpha \in \chi) = P^\dagger(\mathbf{I} = \mathbf{i})P_{\mathbf{i}}^\dagger(\xi_\alpha, \alpha \in \chi_{\mathbf{i}})P_{\mathbf{i}^c|\mathbf{i}}(\xi_\alpha, \alpha \notin \chi_{\mathbf{i}}|\xi_\alpha, \alpha \in \chi_{\mathbf{i}}) \tag{4.31}$$

where

$$P^\dagger(\mathbf{I} = \mathbf{i}) = \frac{P(\mathbf{I} = \mathbf{i})\mathbb{E}f^2(\mathbf{i}, \Xi_\alpha, \alpha \in \chi_{\mathbf{i}})}{\mathbb{E}f^2(\mathbf{I}, \Xi_\alpha, \alpha \in \chi_{\mathbf{I}})}, \tag{4.32}$$

and

$$P_{\mathbf{i}}^\dagger(\xi_\alpha, \alpha \in \chi_{\mathbf{i}}) = \frac{f^2(\mathbf{i}, \xi_\alpha, \alpha \in \chi_{\mathbf{i}})}{\mathbb{E}f^2(\mathbf{i}, \Xi_\alpha, \alpha \in \chi_{\mathbf{i}})}P_{\mathbf{i}}(\xi_\alpha, \alpha \in \chi_{\mathbf{i}}). \tag{4.33}$$

From this point, we denote $\mathbf{I}^\dagger$ for $\mathbf{I}$ that is generated from (4.32). Notice that the representation of $P_{\mathbf{i}}^\dagger(\xi_\alpha, \alpha \in \chi)$ in (4.31) allows us to construct $\mathbf{I}^\dagger$ and $\{\Xi_\alpha^\dagger, \alpha \in \chi\}$ with distribution $P^\dagger(\mathbf{i}, \xi_\alpha, \alpha \in \chi)$ parallel to $\mathbf{I}$ and $\{\Xi_\alpha, \alpha \in \chi\}$ with distribution $P(\mathbf{i}, \xi_\alpha, \alpha \in \chi)$ in (4.30). That is, we first choose $\mathbf{I}^\dagger$ by $P^\dagger(\mathbf{I} = \mathbf{i})$ and then generate $\{\Xi_\alpha^\dagger, \alpha \in \chi_{\mathbf{i}}\}$ following distribution $P_{\mathbf{i}}^\dagger(\xi_\alpha, \alpha \in \chi_{\mathbf{i}})$ given $\mathbf{I}^\dagger = \mathbf{i}$. Finally, we generate $\{\Xi_\alpha^\dagger, \alpha \notin \chi_{\mathbf{i}}\}$ according to $P_{\mathbf{i}^c|\mathbf{i}}(\xi_\alpha, \alpha \notin \chi_{\mathbf{i}}|\xi_\alpha, \alpha \in \chi_{\mathbf{i}})$. As the last term in (4.30) and (4.31) are exactly the same, the reader will see in the construction below that this equality will allow us to set $\Xi_\alpha = \Xi_\alpha^\dagger$ for most α, yielding that $UY^\dagger + (1 - U)Y^\ddagger$ and Y' are close.

Construction of an approximate zero bias coupling (Y', Y^*):

Now we construct an approximate zero bias variable Y^* from Y', starting with its underlying permutations. Recall again that we consider the symbols π and Y interchageable with π' and Y', respectively. Let π have the Ewens $\mathcal{E}_\theta$ distribution. By replacing $a_{i,j}$ by $a_{i,j} - a_{\bullet,\bullet}$ we may assume that (4.4) is satisfied, that is, that $\mathbb{E}Y' = 0$ and $a_{\bullet,\bullet} = 0$ and thus $M = \max_{i,j} |a_{i,j}|$.

Now we follow the outline described right before this construction to produce a coupling of Y' to a pair $Y^\dagger, Y^\ddagger$, with the square bias distribution (2.6) and then construct a coupling of Y' to Y^* satisfying (2.3), using uniform interpolation as in Lemma 2.1.

We first provide a brief overview of the construction, providing the full details later. To specialize the outline above to the case at hand, we let $f = b$ with b given in Lemma 4.5, $\mathbf{I} = (I, J)$, $\mathbf{i} = (i, j)$, $\Xi_\alpha = \pi(\alpha)$, $\Xi_\alpha^\dagger = \pi^\dagger(\alpha)$, $\chi = [n]$ and $\xi_\alpha, \alpha \in [n]$ denote distinct elements of $[n]$. As in Lemma 4.5, we write $\pi(-\alpha) = \pi^{-1}(\alpha)$ for $\alpha \in [n]$ and let $\chi_\mathbf{i} = \chi_{i,j} = \{-i, -j, i, j\}$ and

$$p_{i,j}(\xi_\alpha, \alpha \in \chi_{i,j}) = P(\pi(\alpha) = \xi_\alpha, \alpha \in \chi_{i,j}),$$

the distribution of the pre and post images of i and j under π.

By the decomposition in (4.30) and Lemma 4.6, π' and π'' can be constructed by choosing $\mathbf{I}$ with $P(\mathbf{I} = \mathbf{i})$ uniformly on $[n]$, then constructing the pre and post images of I and J under π' and the values of π' on the remaining variables conditional on what has already been chosen and finally letting $\pi'' = \tau_{I,J}\pi'\tau_{I,J}$.

For the distribution of the pair $(Y^\dagger, Y^\ddagger)$ with the square bias distribution, we will first construct the underlying permutations $(\pi^\dagger, \pi^\ddagger)$. For this purpose, we follow the parallel decomposition of $P^\dagger(\mathbf{i}, \xi_\alpha, \alpha \in \chi)$ in (4.31) beginning with the indices $\mathbf{I}^\dagger = (I^\dagger, J^\dagger)$ with distribution (4.32),

$$P(I^\dagger = i, J^\dagger = j) = \frac{P(I = i, J = j)\mathbb{E}b^2(i, j, \pi(\alpha), \alpha \in \chi_{i,j})}{\mathbb{E}b^2(I, J, \pi(\alpha), \alpha \in \chi_{I,J})} \tag{4.34}$$

where $b(i, j, \pi(\alpha), \alpha \in \chi_{i,j})$ is given in Lemma 4.5 and $\mathbb{E}b^2(I, J, \pi(\alpha), \alpha \in \chi_{I,J}) = \mathbb{E}[(Y' - Y'')^2]$ has been calculated in Lemma 4.8. Next, given $I^\dagger = i$ and $J^\dagger = j$, we generate their pre and post images $\pi^{-\dagger}(I^\dagger), \pi^{-\dagger}(J^\dagger), \pi^\dagger(I^\dagger), \pi^\dagger(J^\dagger)$ with distribution (4.33),

$$p_{i,j}^\dagger(\xi_\alpha, \alpha \in \chi_{i,j}) = \frac{b^2(i, j, \xi_\alpha, \alpha \in \chi_{i,j})}{\mathbb{E}b^2(i, j, \pi(\alpha), \alpha \in \chi_{i,j})}p_{i,j}(\xi_\alpha, \alpha \in \chi_{i,j}). \tag{4.35}$$

Next we will construct the remaining images of $\pi^\dagger$ from π conditional on what has already been chosen so that $\mathbf{I}^\dagger$ and $\pi^\dagger$ follow (4.31), and that π and $\pi^\dagger$ are close. Specializing the last factor of (4.31) to the case at hand, the remaining images of $\pi^\dagger$ given what has already been generated have distribution

$$p_{\mathbf{i}^c|\mathbf{i}}^\dagger(\xi_\alpha, \alpha \notin \chi_{i,j}|\xi_\alpha, \alpha \in \chi_{i,j})$$
$$= P\left(\pi^\dagger(\alpha) = \xi_\alpha, \alpha \notin \chi_{i,j}|\pi^\dagger(\alpha) = \xi_\alpha, \alpha \in \chi_{i,j}\right) \tag{4.36}$$

with P follows the original Ewens $\mathcal{E}_\theta$ distribution. As the approximate λ, R-Stein pair (Y', Y'') is defined as in (4.1) with π replaced by $\pi'' = \tau_{I,J}\pi'\tau_{I,J}$ for Y'', we will then let $Y^\dagger$ and $Y^\ddagger$ be as in (4.1) with π replaced by $\pi^\dagger$ and $\pi^\ddagger = \tau_{I^\dagger,J^\dagger}\pi^\dagger\tau_{I^\dagger,J^\dagger}$, respectively. Applying Lemma 2.3, $(Y^\dagger, Y^\ddagger)$ will have the square bias distribution as in (2.6).

Next we construct $\pi^\dagger$ from the given π'. Recall from Section 3 that, with $B \subset [n]$, the reduced permutation $\sigma \backslash B$ acts on $[n] \backslash B$ and has cycle representation obtained by deleting all elements of B in the cycle representation of σ. In view of Proposition 3.2, we will keep the number of cycles $\#(\pi^\dagger \backslash \chi_{i,j})$ of the reduced permutation $\pi^\dagger \backslash \chi_{i,j}$, the same as $\#(\pi' \backslash \chi_{i,j})$ when applying (4.36).

For ease of notation, we denote the values

$$I^\dagger, J^\dagger, \pi^{-\dagger}(I^\dagger), \pi^{-\dagger}(J^\dagger), \pi^\dagger(I^\dagger), \pi^\dagger(J^\dagger) \tag{4.37}$$

generated in (4.34) and (4.35) by i, j, r, s, k, l, respectively. By (4.34), $i \neq j$ and thus $r \neq s$ and $k \neq l$. Lemma 4.4 gives that, for $\pi^\dagger$, membership in $A_0, \ldots, A_4$ and $A_{5,1}, \ldots, A_{5,4}$, defined in Lemma 4.4, is determined by the generated i, j, r, s, k, l.

The remaining specification of $\pi^\dagger$ from π depends on which case, or subcase, of the events $A_0, A_1, A_2, A_3, A_4, A_5$ is obtained. In each instance, we construct $\pi^\dagger$ from π by removing i, j, r, s from π's cycle representation and inserting them into the resulting reduced permutation so that, respectively, r and k will be the pre and post images of i, and s and l will be those of j. As we only delete i, j, r, s from π and then insert these values without moving the remaining ones when constructing $\pi^\dagger$, it is clear that $\pi^\dagger \setminus \{i, j, r, s\}$ and $\pi \setminus \{i, j, r, s\}$ are identical. As the distribution of $\mathbf{I}^\dagger$ and their pre and post images follow (4.32) and (4.33), respectively, and the conditional distribution (4.36) of the remaining values is the same as that for those of π, the pair $(\mathbf{I}^\dagger, \pi^\dagger)$ follows the distribution in (4.31).

In the following, we explicitly construct $\pi^\dagger$ from π by separating into cases, or subcases, of the events $A_0, A_1, A_2, A_3, A_4, A_5$. To be more easily understandable, we also provide an example in Figure 4.1. Although there are several cases, the idea in each case follows the same rule explained in the previous paragraph. Therefore, we only present the full detail in the first nonzero case A_1 and remark for the reader that the last case $A_{5,4}$ is the only one that contributes to the orders of the bounds in Theorem 4.1 and hence it suffices for the reader to read only the first nonzero and the last cases.

Below, we use the notation $i \to j$ if we change the cycle structure of π so that $\pi^\dagger(i) = j$. We say 'delete i from π' if we delete i from the cycle structure of π and connect $\pi^{-1}(i) \to \pi(i)$ so that we end up with the reduced permutation $\pi \setminus \{i\}$. We say 'insert i in front of k' if we put i between $\pi^{-1}(k)$ and k, that is, we end up with $\pi^{-1}(k) \to i \to k$ after the insertion.

Case A_0: As $b_0 = 0$ by Lemma 4.5, $p_{i,j}^\dagger$ as in (4.35) in the case A_0 is zero and therefore we need not consider this case.

Case A_1: We separate this case into two subcases, $A_{1,1} = \{|i| = 1, |j| = 2\}$ and $A_{1,2} = \{|i| = 1, |j| \geq 3\}$. For $A_{1,1}$, we have $(i, j, r, s, k, l) = (i, j, i, s, i, s)$. We first recall that from (4.37) r is the pre image of i and thus $r = i$ means i must be a fixed point. Similarly, since s and l are pre and post images of j, respectively, $l = s \neq j$ implies that (s, j) must be a 2-cycle. Hence, in this case, we simply delete i, j, s from π, then let i be a 1-cycle and (s, j) be a 2-cycle. For $A_{1,2}$, we have $(i, j, r, s, k, l) = (i, j, i, s, i, l)$. By the same reasoning as for $A_{1,1}$, we keep l at the original place and delete i, j, s from π, then let i be a 1-cycle and insert $s \to j$ in front of l. For the remaining unmentioned values of the permutation, we keep them at the original places. Now we call the modified permutation $\pi^\dagger$. Since we have moved only i, j and s, the reduced permutations $\pi^\dagger \setminus \{i, j, s\}$ and $\pi \setminus \{i, j, s\}$ are exactly the same. Applying Proposition 3.2 with $B = \{i, j, s\}$, we have that the remaining images of $\pi^\dagger$, conditioning on $\{\pi^\dagger(i) = i, \pi^\dagger(j) = l, \pi^\dagger(s) = j\}$ has distribution $p_{\mathbf{i}^c|\mathbf{i}}^\dagger(\xi_\alpha, \alpha \notin \{-i, -j, i, j\} | \xi_{-i} = i, \xi_{-j} = s, \xi_i = i, \xi_j = l)$ in (4.36). Therefore the distribution of $(\mathbf{I}^\dagger, \pi^\dagger)$ in A_1 case follows (4.31) and we have

$$\pi(\phi) = \pi^\dagger(\phi) \quad \text{for all} \quad \phi \notin \mathcal{I}_1$$
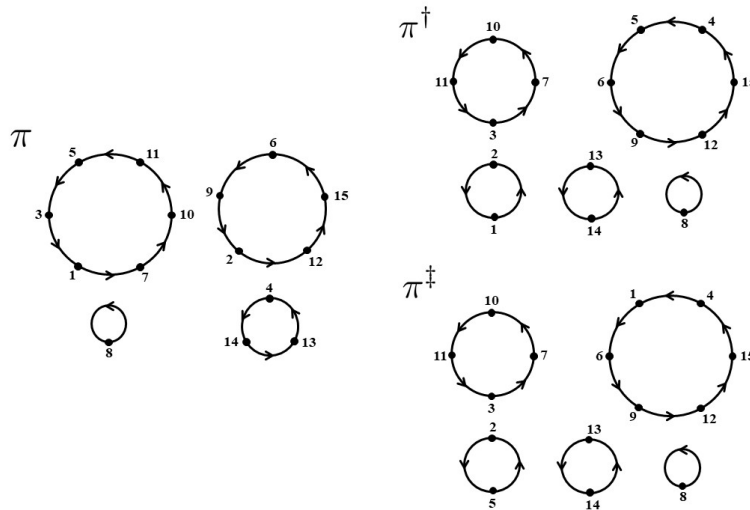$$\text{where} \quad \mathcal{I}_1 = \{\pi^{-1}(\alpha), \beta : \alpha \in \{i, j, s, l\}, \beta \in \{i, j, s\}\}.$$

FIGURE 4.1. This figure shows an example of the construction of $\pi^\dagger$ and $\pi^\ddagger$ from π when $n = 15$ and $(i, j, r, s, k, l) = (1, 5, 2, 4, 2, 6)$ that corresponds to the case $A_{5,2}$. One can easily notice that the reduced permutations $\pi \setminus \{1, 5, 2, 4\}$, $\pi^\dagger \setminus \{1, 5, 2, 4\}$, and $\pi^\ddagger \setminus \{1, 5, 2, 4\}$ have the same cycle structure, that is, one 1-cycle, one 2-cycle and two 4-cycles.

Hence $|\mathcal{I}_1|$ is at most 7. From this point, for any index γ, we denote $\mathcal{I}_\gamma$ for the set that $\pi(\phi) = \pi^\dagger(\phi)$ for all $\phi \notin \mathcal{I}_\gamma$ in the case A_γ.

Case A_2: Switching the role of i and j, we follow the same construction as for A_1 and thus $|\mathcal{I}_2|$ is also at most 7.

Case A_3: Again we separate this case into $A_{3,1} = \{|i| = 3, \pi^\dagger(i) = j\}$ and $A_{3,2} = \{|i| \geq 4, \pi^\dagger(i) = j\}$. For $A_{3,1}$, $(i, j, r, s, k, l) = (i, j, r, i, j, r)$, we delete i, j, r from π and let them be a 3-cycle $r \to i \to j \to r$. For $A_{3,2}$, $(i, j, r, s, k, l) = (i, j, r, i, j, l)$, we delete i, j, r from π and insert $r \to i \to j$ in front of l. By the same reasoning as in A_1, the remaining images of $\pi^\dagger$, conditioning on $\{\pi^\dagger(i) = j, \pi^\dagger(j) = l, \pi^\dagger(r) = i\}$ has distribution $p^\dagger_{\mathbf{i}^c|\mathbf{i}}(\xi_\alpha, \alpha \notin \{-i, -j, i, j\}|\xi_{-i} = r, \xi_{-j} = i, \xi_i = j, \xi_j = l)$ in (4.36) and thus the distribution of $(\mathbf{I}^\dagger, \pi^\dagger)$ in A_3 case follows (4.31). In this case,

$$\pi(\phi) = \pi^\dagger(\phi) \quad \text{for all} \quad \phi \notin \mathcal{I}_3$$
$$\text{where} \quad \mathcal{I}_3 = \{\pi^{-1}(\alpha), \beta : \alpha \in \{i, j, r, l\}, \beta \in \{i, j, r\}\}.$$

Hence $|\mathcal{I}_3|$ is at most 7.

Case A_4: Switching the role of i and j, we follow the same construction as for A_3 and so $|\mathcal{I}_4|$ is at most 7.

Moving on to A_5, here we separate it into $A_{5,1}$, $A_{5,2}$, $A_{5,3}$ and $A_{5,4}$, defined in Lemma 4.4.

Case $A_{5,1}$: Here $(i, j, r, s, k, l) = (i, j, r, s, r, s)$, we delete i, j, r, s from π and then let both (r, i) and (s, j) be 2-cycles. By the same argument as in the previous

cases, the distribution of $(\mathbf{I}^\dagger, \pi^\dagger)$ follows (4.31) and we end up with

$$\pi(\phi) = \pi^\dagger(\phi) \ \text{ for all } \ \phi \notin \mathcal{I}_{5,1}$$
$$\text{where } \ \mathcal{I}_{5,1} = \{\pi^{-1}(\alpha), \alpha : \alpha \in \{i, j, r, s\}\}.$$

Thus $|\mathcal{I}_{5,1}|$ is at most 8.

Case $\mathbf{A_{5,2}}$: As $(i, j, r, s, k, l) = (i, j, r, s, r, l)$, we delete i, j, r, s from π and then let (r, i) be a 2-cycle and insert $s \to j$ in front of l. In this case, again the distribution of $(\mathbf{I}^\dagger, \pi^\dagger)$ follows (4.31) and

$$\pi(\phi) = \pi^\dagger(\phi) \ \text{ for all } \ \phi \notin \mathcal{I}_{5,2}$$
$$\text{where } \ \mathcal{I}_{5,2} = \{\pi^{-1}(\alpha), \beta : \alpha \in \{i, j, r, s, l\}, \beta \in \{i, j, r, s\}\}.$$

Hence $|\mathcal{I}_{5,2}|$ is at most 9.

Case $\mathbf{A_{5,3}}$: Switching the role of i and j, we follow the same construction as for $A_{5,2}$ and thus $|\mathcal{I}_{5,3}|$ is at most 9.

Case $\mathbf{A_{5,4}}$: We again need to break this case into subcases depending on the size of $\{i, j, r, s, k, l\}$. As the case that contributes the most to the difference between π and $\pi^\dagger$ is the case where i, j, r, s, k, l are all distinct, we only discuss the construction of this particular case. We delete i, j, r, s from π and then insert $r \to i$ and $s \to j$ in front of k and l, respectively. In this case, the distribution of $(\mathbf{I}^\dagger, \pi^\dagger)$ again follows (4.31) and

$$\pi(\phi) = \pi^\dagger(\phi) \ \text{ for all } \ \phi \notin \mathcal{I}_{5,4}$$
$$\text{where } \ \mathcal{I}_{5,4} = \{\pi^{-1}(\alpha), \beta : \alpha \in \{i, j, r, s, k, l\}, \beta \in \{i, j, r, s\}\}.$$

Thus $|\mathcal{I}_{5,4}|$ is at most 10.

Now the construction of $\pi^\dagger$ has been specified in every case and subcase and the distribution of $(\mathbf{I}^\dagger, \pi^\dagger)$ follows (4.31). Therefore, setting

$$\pi^\ddagger = \tau_{I^\dagger, J^\dagger} \pi^\dagger \tau_{I^\dagger, J^\dagger}$$

results in a collection of variables $I^\dagger, J^\dagger$ and a pair of permutations with the square bias distribution (2.5). Hence, letting $Y', Y^\dagger, Y^\ddagger$ given by (4.1) with $\pi', \pi^\dagger$ and $\pi^\ddagger$, respectively, yields a coupling of Y' to the variables $Y^\dagger, Y^\ddagger$ with the square bias distribution as required in Lemma 2.1. Invoking Lemma 2.1 to $Y^* = UY^\dagger + (1 - U)Y^\ddagger$ with $U \sim \mathcal{U}[0, 1]$ be independent of $Y^\dagger, Y^\ddagger$, we have Y^* be an approximate Y'-zero bias variable satisfying (2.3), as desired.

The following lemma provides a bound of the difference between Y^* and Y'.

Lemma 4.10. *Let Y' and M be defined as in the statement of Theorem 4.1 and Y^* be constructed as in the construction above. Then $|Y^* - Y'| \leq 20M$.*

Proof: By the same argument as in the proof of Lemma 6.10 in Chen et al. (2011), we have

$$|Y^* - Y'| \leq 2 \max_{\gamma \in \Gamma} |\mathcal{I}_\gamma| M,$$

where $\Gamma = \{1, 2, 3, 4, (5, 1), (5, 2), (5, 3), (5, 4)\}$. Then the claim follows from that $|\mathcal{I}_\gamma|$ is at most 10 from the construction above.

$\square$

Now we have all ingredients to prove our main theorem.

Proof of Theorem 4.1 As before, by replacing $a_{i,j}$ by $a_{i,j} - a_{\bullet,\bullet}$ we may assume that (4.4) is satisfied, that is, that $\mathbb{E}Y' = 0$ and $a_{\bullet,\bullet} = 0$ and thus proceed the construction with $M = \max_{i,j} |a_{i,j}|$. First we construct an approximate $4/n, R$-Stein pair (Y', Y'') as in Lemma 4.6 with the remainder R given in (4.13). Then we construct an approximate zero bias variable Y^* satisfying (2.3) as in the construction right above this proof.

Now we apply Theorems 2.4 and 2.5, handling three terms on the right hand side of (2.7) and (2.8). We note that the three terms from the two theorems are different only on their constants so we handle both L^1 and L^∞ upper bounds at the same time. For the first term, by Lemma 4.10, we have

$$\frac{1}{\sigma}|Y^* - Y'| \leq \frac{20M}{\sigma}. \tag{4.38}$$

Now we handle the last two terms containing the remainder R. Applying (4.17) and (4.16) in Lemma 4.7, respectively, and using that $\lambda = 4/n$ and $n \geq 6$, we obtain

$$\frac{|\mathbb{E}Y'R|}{\sigma^2\lambda} \leq \left(3\kappa_{\theta,n,1} + 1.2\theta + \frac{\kappa_{\theta,n,1}(\theta+1) + \kappa_{\theta,n,2}}{n-1}\right)\frac{M}{\sigma}, \tag{4.39}$$

and

$$\frac{\mathbb{E}|R|}{\sigma\lambda} \leq \frac{0.6M\theta(6n + 4\theta - 5)}{\sigma(\theta + n - 1)} + \frac{0.5\theta^2 Mn}{\sigma(\theta + n - 1)_{(2)}}, \tag{4.40}$$

where $\kappa_{\theta,n,1}$ and $\kappa_{\theta,n,2}$ are given in (4.7) and (4.8), respectively.

Invoking Theorems 2.4 and 2.5, using (4.38), (4.39) and (4.40), now yields the L^1 and L^∞ upper bounds in (4.5) and (4.6), respectively.

Next we follow the idea in Englund (1981) to prove the L^∞ lower bound in (4.9) in the case that $a_{i,j}$ are all integers. By Chebyshev's inequality,

$$P(\mathbb{E}Y - \sqrt{3}\sigma < Y < \mathbb{E}Y + \sqrt{3}\sigma) \geq 2/3.$$

The random variable Y is discrete having at most $n!$ possibilities and for each pair of possibilities, the difference between them is at least 1 as $a_{i,j}$ are all integers. Therefore the interval $(\mathbb{E}Y - \sqrt{3}\sigma, \mathbb{E}Y + \sqrt{3}\sigma)$ contains at most $\lceil(2\sqrt{3}\sigma + 1)\rceil$ possible values of Y which implies that the largest point mass $p(n)$ in the distribution of Y satisfies

$$p(n) \geq \frac{2}{6\sqrt{3}\sigma + 3}.$$

As $\Phi(x)$ the distribution function of Z is continuous, we have

$$\sup_{\infty < x < \infty} |P(Y < x) - \Phi((x - \mathbb{E}Y)/\sigma)| \geq \frac{1}{2}p(n) = \frac{1}{6\sqrt{3}\sigma + 3}.$$

$\square$

5. Appendix

We prove Lemmas 4.6 and 4.8 in this section.

Proof of Lemma 4.6 As Ewens measure is constant on cycle type, the exchangeability claim follows immediately from the proof of Lemma 6.9 of Chen et al. (2011).

It remains to show that Y', Y'' satisfies (1.9) with R given by (4.13). As Y' is a function of π' the tower property of conditional expectation yields that

$$\mathbb{E}[Y' - Y''|Y'] = \mathbb{E}[\mathbb{E}[Y' - Y''|\pi]|Y'],$$

and we begin by computing the conditional expectation given π of the difference $\sum_{m=0}^{5} b_m(i,j,\pi(\alpha), \alpha \in \chi_{i,j})\mathbf{1}_{A_m}$ in (4.11) with $A_0, \ldots, A_5$ given in Lemma 4.4, with i,j replaced by I, J.

First we have that $b_0 = 0$. As given in (6.85) of Chen et al. (2011), the contribution to $n(n-1)\mathbb{E}[Y' - Y''|\pi]$ from b_1 and b_2 totals to

$$2(n - c_1) \sum_{|i|=1} a_{i,i} + 4c_1 \sum_{|i|\geq 2} a_{i,\pi(i)} - 2c_1 \sum_{|i|\geq 2} a_{i,i}$$
$$- 2 \sum_{|i|=1,|j|\geq 2} a_{i,j} - 2 \sum_{|i|\geq 2,|j|=1} a_{i,j}, \tag{5.1}$$

and likewise (6.88) shows b_3 and b_4 contribute

$$6 \sum_{|i|\geq 3} a_{i,\pi(i)} - 4 \sum_{|i|\geq 3} a_{\pi^{-1}(i),\pi(i)} - 2 \sum_{|i|\geq 3} a_{\pi(i),i}, \tag{5.2}$$

and (6.90) shows the first four terms of b_5 contribute

$$4(n - 2 - c_1) \sum_{|i|=2} a_{i,\pi(i)} + 4(n - 3 - c_1) \sum_{|i|\geq 3} a_{i,\pi(i)}. \tag{5.3}$$

Lastly, the contribution from the fifth term of b_5 is given by (6.91), and separating out the cases where $i \neq j$ and $i = j$ in the first sum there, that expression can be seen equivalent to

$$-\sum_{|i|\geq 4} \sum_{j\sim i,j\neq i} a_{i,j} + \sum_{|i|\geq 4}(a_{i,\pi(i)} + a_{\pi^{-1}(i),\pi(i)}) - \sum_{|i|\geq 2,|j|\geq 2} \sum_{j\nsim i} a_{i,j}. \tag{5.4}$$

To simplify (5.4), let $a \wedge b = \min(a,b)$, and follow (6.92) of Chen et al. (2011), separating out the cases where $j = i$ and $j \neq i$, resulting in the identity

$$\theta \sum_{i=1}^{n} a_{i,i} + \sum_{i\neq j} a_{i,j} = \theta \sum_{|i|\geq 1} a_{i,i} + \sum_{|i|\geq 4} \sum_{j\sim i,j\neq i} a_{i,j} + \sum_{|i|\leq 3} \sum_{j\sim i,j\neq i} a_{i,j}$$
$$+ \sum_{|i|\geq 2,|j|\geq 2} \sum_{j\nsim i} a_{i,j} + \sum_{|i|\wedge|j|=1} \sum_{j\nsim i} a_{i,j}. \tag{5.5}$$

Since $\theta \sum_{i=1}^{n} a_{i,i} + \sum_{i\neq j} a_{i,j} = n(\theta + n - 1)a_{\bullet,\bullet} = 0$ by the assumption, we replace the sum of the first and last terms in (5.4) by the sum of the first, third and fifth terms in (5.5). Now (5.4) equals

$$\theta \sum_{|i|\geq 1} a_{i,i} + \sum_{|i|\leq 3} \sum_{j\sim i,j\neq i} a_{i,j} + \sum_{|i|\geq 4}(a_{i,\pi(i)} + a_{\pi^{-1}(i),\pi(i)}) + \sum_{|i|\wedge|j|=1} \sum_{j\nsim i} a_{i,j}.$$

Using that $\pi^2(i) = \pi^{-1}(i)$ when $|i| = 3$ and there are only two points in the cycle when $|i| = 2$, we obtain

$$\theta \sum_{|i|\geq 1} a_{i,i} + \sum_{|i|\geq 2} a_{i,\pi(i)} + \sum_{|i|\geq 3} a_{\pi^{-1}(i),\pi(i)} + \sum_{|i|\wedge|j|=1} \sum_{j\nsim i} a_{i,j}.$$

Combining this contribution with the next three terms of b_5, each of which yields the same amount, gives the total

$$4\theta \sum_{|i|\geq 1} a_{i,i} + 4 \sum_{|i|\geq 2} a_{i,\pi(i)} + 4 \sum_{|i|\geq 3} a_{\pi^{-1}(i),\pi(i)} + 4 \sum_{|i|\wedge|j|=1} \sum_{j\nsim i} a_{i,j}. \tag{5.6}$$

Combining (5.6) with the contribution (5.3) from the first four terms in b_5, the b_1 and b_2 terms in (5.1) and the b_3 and b_4 terms (5.2), yields

$$
\begin{aligned}
&n(n-1)\mathbb{E}[Y'-Y''|\pi]\\
&= 4(n-1)\sum_{|i|=2}a_{i,\pi(i)} + (4n-2)\sum_{|i|\geq 3}a_{i,\pi(i)} - 2\sum_{|i|\geq 3}a_{\pi(i),i}\\
&\quad + 2(n-c_1+2\theta)\sum_{|i|=1}a_{i,i} - 2(c_1-2\theta)\sum_{|i|\geq 2}a_{i,i}\\
&\quad + 4\sum_{|i|\wedge|j|=1,j\not\sim i}a_{i,j} - 2\sum_{|i|=1,|j|\geq 2}a_{i,j} - 2\sum_{|i|\geq 2,|j|=1}a_{i,j}\\
&= 4(n-1)\sum_{|i|=2}a_{i,\pi(i)} + (4n-2)\sum_{|i|\geq 3}a_{i,\pi(i)} - 2\sum_{|i|\geq 3}a_{\pi(i),i} + 4(n-1)\sum_{|i|=1}a_{i,i}\\
&\quad - 2(n+c_1-2(\theta+1))\sum_{|i|=1}a_{i,i} - 2(c_1-2\theta)\sum_{|i|\geq 2}a_{i,i}\\
&\quad + 4\sum_{|i|\wedge|j|=1,j\not\sim i}a_{i,j} - 2\sum_{|i|=1,|j|\geq 2}a_{i,j} - 2\sum_{|i|\geq 2,|j|=1}a_{i,j},
\end{aligned}
$$

where the final two expressions are identical but for a rewriting of the coefficient of $\sum_{|i|=1}a_{i,i}$.

Now applying the assumption that $a_{i,j} = a_{j,i}$ to the third term, using that $\pi(i) = i$ for $|i| = 1$ for the fourth term, and $a_{i,j} = a_{j,i}$ again to obtain

$$
\begin{aligned}
4\sum_{|i|\wedge|j|=1,j\not\sim i}a_{i,j} - 2\sum_{|i|=1,|j|\geq 2}a_{i,j} - 2\sum_{|i|\geq 2,|j|=1}a_{i,j}\\
= 4\sum_{|i|=1,|j|=1,j\neq i}a_{i,j} + 4\sum_{|i|=1,|j|\geq 2}a_{i,j}
\end{aligned}
$$

we have

$$
\begin{aligned}
n(n-1)\mathbb{E}[Y'-Y''|\pi] &= 4(n-1)\sum_{|i|=2}a_{i,\pi(i)} + (4n-2)\sum_{|i|\geq 3}a_{i,\pi(i)}\\
&\quad -2\sum_{|i|\geq 3}a_{i,\pi(i)} + 4(n-1)\sum_{|i|=1}a_{i,\pi(i)} - T\\
&= 4(n-1)Y' - T,
\end{aligned}
$$

where T is given in (4.14). Now conditioning on Y' proves the claim that $\mathbb{E}[Y''|Y'] = (1-\lambda)Y' + R(Y')$ with $\lambda = 4/n$ and $R(Y')$ as in (4.13).

$\square$

In the following proof, for ease of notation, we write $\Sigma_{\alpha_1,\ldots,\alpha_m}$ for the sum over distinct $\alpha_k \in [n]$.

Proof of Lemma 4.8 We first calculate $\mathbb{E}[(Y'-Y'')^2]$ with the help of Lemma 4.5. Note that we write $b_m(i,j,\pi(\alpha), \alpha \in \chi_{i,j}) = b_m(i,j,\pi^{-1}(i),\pi^{-1}(j),\pi(i),\pi(j))$ in what follows. As $b_0 = 0$, moving on to A_1, and recalling that the summations in

this proof are over distinct values, we have

$$\mathbb{E}[(Y'-Y'')^2]\mathbf{1}_{A_1} = \mathbb{E}b_1^2(I,J,\pi^{-1}(I),\pi^{-1}(J),\pi(I),\pi(J))\mathbf{1}_{\{\pi(I)=I,\pi(J)\neq J\}}$$

$$= \frac{1}{n(n-1)(\theta+n-1)_{(3)}}\left(\theta^2\sum_{i,j,s}b_1^2(i,j,i,s,i,s)+\theta\sum_{i,j,s,l}b_1^2(i,j,i,s,i,l)\right),$$

noting that there are $n(n-1)$ possibilities for I and J, applying Proposition 3.1 with $B=\{i,j,s\}$ we see that the factor $\theta^2/(\theta+n-1)_{(3)}$ in the first term is the probability that $\pi(i)=i$, $\pi(j)=s$ and $\pi(s)=j$ and the factor $\theta/(\theta+n-1)_{(3)}$ in the second term is the probability that $\pi(i)=i$, $\pi(j)=l$ and $\pi(s)=j$. By symmetry, A_2 contributes the same amount.

For A_3, by similar reasoning, we have,

$$\mathbb{E}[(Y'-Y'')^2]\mathbf{1}_{A_3} = \mathbb{E}b_3^2(I,J,\pi^{-1}(I),\pi^{-1}(J),\pi(I),\pi(J))\mathbf{1}_{\{\pi(I)=J,\pi(J)\neq I\}}$$

$$= \frac{1}{n(n-1)(\theta+n-1)_{(3)}}\left(\theta\sum_{i,j,r}b_3^2(i,j,r,i,j,r)+\sum_{i,j,r,l}b_3^2(i,j,r,i,j,l)\right).$$

The contribution from A_4 is the same as that from A_3.

Lastly, consider the contribution from A_5. Starting with $A_{5,1}$, we have

$$\mathbb{E}[(Y'-Y'')^2]\mathbf{1}_{A_{5,1}} = \mathbb{E}b_5^2(I,J,\pi^{-1}(I),\pi^{-1}(J),\pi(I),\pi(J))\mathbf{1}_{\{|I|=2,|J|=2,I\not\sim J\}}$$

$$= \frac{\theta^2}{n(n-1)(\theta+n-1)_{(4)}}\sum_{i,j,r,s}b_5^2(i,j,r,s,r,s).$$

For $A_{5,2}$, we have

$$\mathbb{E}[(Y'-Y'')^2]\mathbf{1}_{A_{5,2}} = \mathbb{E}b_5^2(I,J,\pi^{-1}(I),\pi^{-1}(J),\pi(I),\pi(J))\mathbf{1}_{\{|I|=2,|J|\geq 3\}}$$

$$= \frac{\theta}{n(n-1)(\theta+n-1)_{(4)}}\sum_{i,j,r,s,l}b_5^2(i,j,r,s,r,l).$$

Again by symmetry, $A_{5,3}$ contributes the same.

For $A_{5,4}$, we have

$$\mathbb{E}[(Y'-Y'')^2]\mathbf{1}_{A_{5,4}} = \mathbb{E}b_5^2(I,J,\pi^{-1}(I),\pi^{-1}(J),\pi(I),\pi(J))\mathbf{1}_{\{|I|\geq 3,|J|\geq 3\}}$$

$$= \frac{1}{n(n-1)(\theta+n-1)_{(4)}}\left(\theta\sum_{i,j,r,k}b_5^2(i,j,r,k,k,r)\right.$$

$$\left.+\sum_{i,j,r,k,l}b_5^2(i,j,r,k,k,l)+\sum_{i,j,r,s,k,l}b_5^2(i,j,r,s,k,l)\right).$$

Summing all cases, yields $\mathbb{E}[(Y'-Y'')^2]$ in (4.22).

Since $\mathbb{E}[(Y'-Y'')^2]=2\left(\lambda\sigma^2-\mathbb{E}Y'R\right)$ and $\lambda=4/n$, we have

$$\sigma^2 = \frac{n\mathbb{E}[(Y'-Y'')^2]}{8}+\frac{n\mathbb{E}Y'R}{4}$$

$$= \frac{n}{8}(2\beta_1+2\beta_3+\beta_{5,1}+2\beta_{5,2}+\beta_{5,4})+\frac{n\mathbb{E}Y'R}{4},$$

where β_1, β_3, $\beta_{5,1}$, $\beta_{5,2}$ and $\beta_{5,4}$ are defined as in (4.24), (4.25), (4.26), (4.27) and (4.28), respectively.

$\square$

Acknowledgements

References

D. J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin (1985). MR883646.

R. Arratia, A. D. Barbour and S. Tavaré. *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich (2003). ISBN 3-03719-000-0. MR2032426.

L. H. Y. Chen and X. Fang. On the error bound in a combinatorial central limit theorem. *Bernoulli* **21** (1), 335–359 (2015). MR3322321.

L. H. Y. Chen, L. Goldstein and Q.-M. Shao. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg (2011). ISBN 978-3-642-15006-7. MR2732624.

G. Englund. A remainder term estimate for the normal approximation in classical occupancy. *Ann. Probab.* **9** (4), 684–692 (1981). MR624696.

W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Population Biology* **3**, 87–112; erratum, ibid. 3 (1972), 240; erratum, ibid. 3 (1972), 376 (1972). MR0325177.

A. N. Frolov. Esseen type bounds of the remainder in a combinatorial CLT. *J. Statist. Plann. Inference* **149**, 90–97 (2014). MR3199896.

A. N. Frolov. On Esseen type inequalities for combinatorial random sums. *Comm. Statist. Theory Methods* **46** (12), 5932–5940 (2017). MR3620925.

J. Fulman and L. Goldstein. Zero biasing and Jack measures. *Combin. Probab. Comput.* **20** (5), 753–762 (2011). MR2825589.

L. Goldstein. Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42** (3), 661–683 (2005). MR2157512.

L. Goldstein and Ü. Işlak. Concentration inequalities via zero bias couplings. *Statist. Probab. Lett.* **86**, 17–23 (2014). MR3162712.

L. Goldstein and G. Reinert. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** (4), 935–952 (1997). MR1484792.

S. T. Ho and L. H. Y. Chen. An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probability* **6** (2), 231–249 (1978). MR0478291.

W. Hoeffding. A combinatorial central limit theorem. *Ann. Math. Statistics* **22**, 558–566 (1951). MR0044058.

J. Pitman. Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, probability and game theory*, volume 30 of *IMS Lecture Notes Monogr. Ser.*, pages 245–267. Inst. Math. Statist., Hayward, CA (1996). MR1481784.

N. Ross. Fundamentals of Stein's method. *Probab. Surv.* **8**, 210–293 (2011). MR2861132.

C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables pages 583–602 (1972). MR0402873.

G. A. Watterson. The sampling theory of selectively neutral alleles. *Advances in Appl. Probability* **6**, 463–488 (1974). MR0351504.